PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/00 | A2

(11) International Publication Number: **WO 99/60103**

(43) International Publication Date: 25 November 1999 (25.11.99)

(21) International Application Number: PCT/GB99/01610

(22) International Filing Date: 21 May 1999 (21.05.99)

(30) Priority Data:
9810997.8 21 May 1998 (21.05.98) GB

(71) Applicant *(for AE AU BB CA CY GB GD GH GM IE IL KE LC LK LS MN MW NZ SD SG SL SZ TT UG ZA ZW only)*: UNILEVER PLC [GB/GB]; Unilever House, Blackfriars, London EC4P 4BQ (GB).

(71) Applicant *(for all designated States except AE AU BB CA CY GB GD GH GM IE IL IN KE LC LK LS MN MW NZ SD SG SL SZ TT UG US ZA ZW)*: UNILEVER N.V. [NL/NL]; Weena 455, NL–3013 AL Rotterdam (NL).

(71) Applicant *(for IN only)*: HINDUSTAN LEVER LIMITED [IN/IN]; Hindustan Lever House, 165/166 Backbay Reclamation, Mumbai 400 020, Maharashtra (IN).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GIDLEY, Michael, John [GB/GB]; Unilever Research, Colworth Laboratory, Colworth House, Sharnbrook, Bedford MK44 1LQ (GB). CHENGAPPA, Sumant [IN/GB]; Unilever Research, Colworth Laboratory, Colworth House, Sharnbrook, Bedford MK44 1LQ (GB). REID, John, Spence, Grant [GB/GB]; University of Sterling, Dept. of Biological Sciences, Stirling FK9 4LA (GB). EDWARDS, Mary, Elizabeth [GB/GB]; University of Sterling, Dept. of Biological Sciences, Stirling FK9 4LA (GB). DICKSON, Cathryn, Anne [GB/GB]; University of Sterling, Dept. of Biological Sciences, Stirling FK9 4LA (GB).

(74) Agents: KREMER, Simon, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GENES AND ENZYMES

(57) Abstract

Provided are genes encoding enzymes (e.g. galactosyltransferases from fenugreek and guar) which have a role in the biosynthesis of complex non–cellulosic cell wall polysaccharides such as galactomannan. Variants and other products based on the genes are also provided, as are antibodies to the enzymes, plus also methods of isolating or preparing any of these. Also disclosed are vectors and other methods and materials which may be used for cloning the genes (or related nucleic acids e.g. anti–sense versions) into hosts such as transgenic plants having modified polysaccharides. Further disclosed are products, compositions and methods employing these plants and polysaccharides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GENES & ENZYMES

TECHNICAL FIELD

The present invention relates to nucleic acids and other materials having a role in the biosynthesis of complex non-cellulosic plant cell wall polysaccharides. It further relates to various applications for such materials.

RELEVANT PRIOR ART

Plant cell walls contain a number of non-cellulosic polysaccharides which play important roles *in vivo* both structurally and as energy stores. Some of these are discussed in Brett & Waldron (1996) "Physiology and Biochemistry of Plant Cell Walls - 2nd Edition" Pub. Chapman & Hall, London, especially pages 4-43. Generally these occur in the cell wall matrix phase as pectins and hemicelluloses.

Two principal cell wall storage polysaccharides (CWSPs) are the hemicelluloses galactomannan (e.g. guar gum, locust bean gum) and xyloglucan (e.g. tamarind seed polysaccharide). The various known characteristics of these CWSPs, including their structure, application to industry, and metabolism is summarised in Reid & Edwards (1995) "Galactomannans and other cell wall storage polysaccharides in seeds" in "Food polysaccharides and their applications" Ed: Stephen, pp 155-186; Pub. Marcel Dekker.

Role of galactomannans in vivo

Galactomannans are found in the endosperm cells of leguminous seeds, and in the endosperms of the seeds of a small number of non-leguminous species. In general they act as storage reserves, being broken down following germination to monosaccharides which are used by the developing seedling. Their overall biological functions are more complex. The galactomannan of fenugreek has been shown to be multifunctional, imbibing large amounts of water during seed hydration, deploying it as a buffer to protect the germinating embryo from post-imbibition drought, and serving as a substrate reserve following successful germination.

Structure of galactomannans

Structurally galactomannans comprise a (1-4)-β-linked D-mannan backbone which carries single-unit α-D-galactosyl substituents attached (1-6)-α to backbone mannose. Mannose/Galactose [Man/Gal] ratios in galactomannans range from about 3.5 [low-galactose] to about 1.1 [high galactose].

In the legumes, Man/Gal ratio is constant and fixed for the galactomannan of a given species [genetic control], and Man/Gal ratios are similar within taxonomic sub-groupings of the Leguminosae. Species with similar Man/Gal ratios may differ in the statistical distribution of galactose residues along the mannan backbone. Galactomannans are closely related structurally to other cell wall storage polysaccharides of seeds [mannans, glucomannans, galactoglucomannans] and to non-cellulosic polysaccharides of non-seed plant cell walls [glucomannans, galactoglucomannans].

Galactomannan biosynthesis

Galactomannan biosynthesis has been studied using three model leguminous species representative of those forming high-galactose, medium-galactose and low-galactose galactomannans. These are fenugreek [*Trigonella foenum-graecum*, Man/Gal = 1.1], guar [*Cyamopsis tetragonoloba*, Man/Gal = 1.6] and senna [*Senna occidentalis*, Man/Gal = 3.3] (see Edwards et al (1992) Planta 187:67-74; also (1995) Planta 195: 489-495; also (1989) Planta 178: 41-51).

These workers used mixed membrane preparations prepared from endosperms, hand-dissected from developing seeds. The preparations were enzymatically active, catalysing the formation of labelled polysaccharide from GDP-^{14}C-mannose, from GDP-^{14}C-mannose plus unlabelled UDP-galactose [UDP-Gal]and from unlabelled GDP-mannose [GDP-Man] and UDP-^{14}C-galactose. By acid hydrolysis, and in particular the use of pure structure-sensitive galactomannan-hydrolysing enzymes, the polysaccharide products formed from combinations of UDP-Gal and GDP-Man were shown unequivocally to be galactomannans, and the product formed from GDP-Man alone to be (1-4)-β-linked mannan. Thus galactomannan biosynthesis appears to be catalysed by the interaction of two membrane-bound enzymes - a GDP-Man dependent (1-4)-β-D-mannosyltransferase and a specific, UDP-Gal dependent α-D-galactosyltransferase.

The nature of the interaction between the mannan synthase and the galactosyltransferase was also investigated using the membrane preparations. This demonstrated that the mannan synthase can operate independently of the galactosyltransferase, that the galactosyltransferase cannot operate in the absence of simultaneous mannan synthase action and that (1-4)-β-D-mannan preformed at the site of synthesis using the mannan synthase is not accessible as a substrate for the galactosyltransferase. Thus an experimental model for

galactomannan biosynthesis involves stepwise chain-elongation of the mannan chain towards the non-reducing end catalysed by the mannan synthase and transfer of galactose, catalysed by the galactosyltransferase, to a hypothetical galactosyl acceptor mannose residue at or close to the [elongating] non-reducing chain-end.

The regulation of Man/Gal ratio in galactomannan biosynthesis.
In fenugreek, guar and senna the activities of the mannan synthase and the galactosyltransferase in developing endosperms vary *pari passu* with galactomannan deposition, and the relative amounts of the two activities vary little during the period of deposition. In fenugreek and guar [high- and medium-galactose galactomannans] the Man/Gal ratios of the galactomannan present in the endosperm cell walls during galactomannan deposition remain constant at 1.1 and 1.6 respectively. In senna the Man/Gal ratio increases during late seed development from about 2 to 3.3, and this change is accompanied by the appearance and increase of the activity of a galactomannan-active α-galactosidase. Thus in the high and medium-galactose species Man/Gal ratio is determined only by the pathway of biosynthesis. In the low-galactose species the Man/Gal ratio of the primary biosynthetic product is controlled by the biosynthetic process, and the primary biosynthetic product undergoes a post-depositional modification catalysed by a galactomannan-active α-galactosidase.

In vitro galactomannan biosynthesis.
Labelled galactomannans with a range of Man/Gal ratios can be formed in vitro from UDP-Gal and GDP-Man and the membrane-preparations from fenugreek. This is because the rate of mannan-chain elongation in vitro is independent of the rate of galactosyl transfer. Published work suggests that galactosyl transfer depends on the availability of nascent mannan chain as acceptor substrate, and the enzyme system in vitro forms low-galactose galactomannans when saturating GDP-Man and UDP-Gal concentrations are supplied. By retaining UDP-Gal concentrations at saturating and progressively decreasing the rate of mannan chain extension by lowering the GDP-Man concentration, a range of labelled galactomannan products can be obtained with galactose-contents approaching, but not exceeding, those of the primary products of biosynthesis in vivo. The labelled galactomannans can be fragmented, using a pure structure-sensitive endo-(1-4)-β-D-mannanase, to give a series of diagnostic manno-and galactomanno-oligosaccharides, the relative amounts of which can be determined accurately using quantitative digital autoradiography after separation on thin layer

chromatography [TLC] plates. The results of digital autoradiography comprise a structural "fingerprint" of each in vitro galactomannan.

Computer modelling galactomannan biosynthesis.

The experimental model for the interaction of the mannan synthase and galactosyltransferase in galactomannan biosynthesis has been computer modelled with an inbuilt [second-order Markov chain] assumption that the probability of obtaining galactose-substitution at the galactosylacceptor mannose residue is influenced by the existing states of substitution at the nearest and second-nearest neighbour mannose residues only. Also computer modelled is the substrate specificity of the structure-sensitive endo-(1-4)-β-D-mannanase. Thus a computer algorithm is available which when supplied with a set of four numerical probabilities [P_{00}, P_{10}, P_{01}, P_{11}, corresponding to the possible states of substitution at the nearest and second-nearest neighbour mannose residues] will simulate the synthesis of a galactomannan molecule according to the experimental model, and its hydrolysis by the structure-sensitive endo-mannanase, outputting the relative proportions of the diagnostic manno- and galactomanno-oligosaccharides released. This algorithm has been used to process the quantitative endo-mannanase fragmentation data from the labelled in vitro galactomannans from fenugreek, guar and senna, with input of the experimental data and output, for each galactomannan of a set of four probabilities. The results generate the following three statistical statements.

1. The second-order Markov chain assumption built into the computer simulation of the biosynthetic process is adequate.

2. The specificities of the biosynthetic enzyme systems from fenugreek, guar and senna are different, giving different statistical patterns of galactose-substitution along the mannan backbone.

3. For each species the deduced statistical substitution rules define maximum permitted degrees of galactose-substitution which are approached by the degrees of galactose substitution exhibited by the primary products of galactomannan biosynthesis in vivo.

In biochemical terms:

- The (galacto)mannan substrate subsite recognition of the galactosyltransferases from fenugreek, guar and senna must encompass at least three backbone mannosyl residues: the one which is the site

of reaction, and the two preceding ones, towards the reducing end of the chain. Other backbone mannosyl residues may be recognised by the galactosyltransferase, but their states of substitution do not influence greatly the probability of obtaining galactosyl-substitution at the reacting mannosyl residue.

- Galactosyltransferase specificity regulates the distribution of galactose residues along the galactomannan backbone and also sets a maximum limit the degree of galactosyl substitution attainable for the primary product of biosynthesis in each species.

- This limit is achieved in vivo.

Applications for CWSPs to industry

The complex hydrophilic properties of galactomannans underlie also the industrial applications of galactomannans. For example, in the food industry they are used as stabilisers, emulsifiers and in combination with other polysaccharides and proteins to impart more complex rheologies.

The commercial functionality of galactomannans is dependent upon the Man/Gal ratio and, to a lesser extent, the galactose distribution along the mannan backbone. Generally higher Man/Gal ratios are desirable. Of the two principal commercial galactomannans locust bean gum [Man/Gal = 3.5, galactomannan of *Ceratonia siliqua*] is superior to guar gum [Man/Gal = 1.6, galactomannan of *Cyamopsis tetragonoloba*], particularly in mixed polysaccharide interactions.

EP 0 255 153 (Unilever NV/Unilever plc) discusses the use of recombinantly produced guar alpha-galactosidase for providing galactomannans having improved properties.

WO 97/20937 (Danisco) discusses methods of *in vivo* modification of mannose/galactose ratios in galactomannans. The Examples apparently disclose the cloning of a phosphomannose isomerase gene (involved in mannose-6-phosphate generation) from guar, and also the use of senna alpha-galactosidase.

However it is clear from the discussion above that galactosyltransferases are key enzymes in the regulation of galactose distribution along the backbone and in controlling the Man/Gal ratio. Indeed the importance of glycosyltransferases is acknowledged in WO

97/20937 at page 26. However, notwithstanding this, and the extensive research done on their mechanism using impure membrane preparations, no membrane bound transferases involved in the biosynthesis of non-cellulosic plant cell wall polysaccharides have been purified and no cDNA or genomic DNA sequences encoding such transferases have been identified.

The difficulty in isolating such enzymes is discussed briefly in Reid & Edwards (1995) *supra* at page 164 and Brett & Waldron (1996) at page 79. In particular the plant cell wall is an extremely complex structure making it difficult to purify polysaccharide-acting enzymes, or to associate them with the metabolism of any given wall component. The isolation of enzymes which catalyse the biosynthesis of CCWPs is particularly difficult because they are tightly membrane-bound, to Golgi membranes.

An assay for galactosyltransferase activity, in the form of membrane preparations, is disclosed in Edwards et al (1989) Planta 178: 41-51. As described above, in this assay a radiolabelled sugar nucleotide [glycosyldonor] substrate is supplied, the acceptor [nascent mannan] substrate is believed to be formed by the simultaneous operation of an associated mannan synthase. The labelled polysaccharide product is then isolated. Strict controls are necessary to ensure that the "correct" polysaccharide (galactomannan) is assayed.

However this assay is unsuitable for assaying the enzyme in solubilised form. This in turn means it can not readily be used for the identification and therefore purification of the solubilised enzyme (for instance, to a level sufficient to provide sequence data which could be used to isolate corresponding nucleic acids).

Thus it will be seen from the foregoing that the provision of novel nucleic acids and other materials having a role in the biosynthesis of complex non-cellulosic plant cell wall polysaccharides and/or uses thereof would provide a contribution to the art.

DISCLOSURE OF THE INVENTION

The present inventors have used novel techniques to identify and isolate a membrane-bound glycosyltransferase, and encoding nucleic acid, which catalyses the biosynthesis of a complex non-cellulosic plant cell wall polysaccharide.

The glycosyltransferase has demonstrated activity as a galactosyltransferase, involved in the biosynthesis of galactomannan. The polypeptide has a single membrane-spanning α-helix near the N-terminus which appears to serve to anchor the whole polypeptide to a biological membrane. This is the first time that a plant enzyme with activity appropriate for hemicellulose or pectin synthesis has been isolated, and that a nucleic acid sequence has been positively identified as encoding the same.

Briefly, the inventors showed that using the assay described above with detergent "solubilised" fenugreek material, mannan synthase activity was apparently retained at a low level, whilst galactosyltransferase activity was lost completely.

However they established that soluble acceptor molecules, (manno-oligosaccharides) could be used to mimic the nascent mannan chain. Labelled galacto-manno-oligosaccharide products were then carefully purified from other labelled substances and the galactosyl link in these oligosaccharides was shown to be the correct one for a galactosyltransferase involved in galactomannan synthesis.

Owing to the limited amounts of material available (endosperms were hand-dissected from fenugreek seeds about 5mm in diameter at the correct stage of development) only very small-scale purifications of the detergent-solubilised extract could be contemplated. It was found that if isoelectric focussing [IEF] agarose gels were prepared with the solubilising detergent incorporated and detergent-solubilised extract applied as sample, galactosyltransferase activity survived the IEF procedure and was focussed within the gel. Strips from the gel were analysed in parallel for activity and protein content.

Initially galactosyltransferase activity in the narrow strips cut from the IEF gels was assayed by incubating them in the presence of UDP-14C-Gal and a manno-oligosaccharide (usually mannohexaose) and carrying out a quantitative analysis of 14C present in galactosylmannohexaose after the enzyme reaction. Obtaining accurate analysis data required a multi-step procedure involving ion-exchange chromatography, TLC, digital autoradiography and scintillation counting.

Subsequently the inventors determined that low-galactose (and to a lesser extent medium-galactose) galactomannans would also serve as

acceptors for the detergent-solubilised galactosyltransferase. This effect was quite unexpected as earlier studies carried out using the membrane preparations of the prior art suggested that the limited number of acceptor groups available in these substrates would restrict their usefulness. The inventors further established that activity could be assessed *in situ* in the IEF gels. This could be done using commercial agarose IEF gels which contained a blend of agarose and a galactomannan (apparently locust bean gum). This meant that gel strips incubated in the presence of UDP-Gal could be subjected to a simple washing procedure, after which the radioactivity remaining in the gel strips provided a measure of, and a localisation of, galactosyltransferase activity.

Protein distribution within IEF gel strips was determined using two procedures. In the first, the strips were cut into narrow slices, which were soaked in SDS-PAGE sample buffer and placed within individual sample wells of SDS-PAGE gels. In the second, entire strips were soaked in SDS-PAGE sample buffer turned at right angles and applied as sample to SDS-PAGE gels, giving effectively a 2-dimensional gel, the first dimension being the IEF separation carried out in the presence of the solubilising detergent.

By correlating enzyme activity and protein distribution after IEF in this way the inventors were able to identify a small number of potential "candidate" proteins. Further analysis including Western blotting and the use of different solubilising detergents identified a particular protein with molecular weight about 50K. All protein sequence data required for cloning the corresponding cDNA was obtained from the about 50K protein recovered from SDS gels.

Further analysis demonstrated that the fenugreek sequence encoded a 51K protein, with a single hydrophobic membrane-spanning helix near the N-terminal end. This is typical of golgi-bound enzymes.

The sequence apparently shares limited but significant homology with yeast galactosyltransferases, plus also low homologies with yeast mannan synthases and a plant β-mannanase.

Identity was confirmed by cloning the cDNA in-frame into the genome of *Pichia pastoris* methylotrophic yeast, under the control of an alcohol oxidase promoter, and with the yeast α-secretion factor. Two constructs were made, one with the full cDNA sequence, and the other with the

sequence minus the N-terminal membrane-spanning domain, to avoid expressed protein becoming membrane-bound in the yeast. Culture filtrates were assayed for galactosyltransferase activity. Controls contained none, constructs with the full sequence had moderate activity, and constructs with the curtailed sequence gave extremely high activity.

Each nucleic acid encoding a glycosyltransferase provided by the present inventors may be used to manipulate (e.g. galactomannan) synthesis both *in vitro* and *in vivo* thereby allowing galactomannans to be tailored for particular applications. Additionally it can be used, for instance, to alter the cell wall rheology, and hence mechanical properties (e.g. texture) of plant tissues, thereby permitting the production of improved plants and plant products for consumption or industrial use (e.g. fruits, vegetables, timber, paper etc.).

The galactosyltransferase nucleic acid can also be used to prepare novel genes (variants) having altered properties with respect to the wild-type, or alternatively to facilitate the isolation of homologous genes from natural sources.

In the Examples below, the information provided by the novel fenugreek sequence has been used to assist in the isolation of a guar homolog, the activity of which was confirmed using the same assays as those discussed above.

These and other aspects of the present invention will now be discussed in more detail.

According to a first aspect of the present invention there is provided a nucleic acid molecule encoding a polypeptide which is capable of catalysing the biosynthesis of a complex non-cellulosic plant cell wall polysaccharide.

The polysaccharide may be a pectin or a hemicellulose, preferably the latter. Examples of hemicelluloses include xylan, glucomannan, mannan, galactomannan, glucuronoxylan, xyloglucan, callose or arabinogalactan.

The polypeptide is preferably a glycosyltransferase, which is to say that it catalyses, *inter alia*, the addition of monosaccharides (optionally from an activated precursor or donor e.g. a sugar nucleotide, such as a diphosphate precursor e.g. ADP- CDP- GDP- TDP- or

UDP-sugar) to a polysaccharide chain (the 'acceptor') generally, but not exclusively, at the non-reducing end. Such enzymes are occasionally also termed polysaccharide synthases or synthetases by those skilled in the art.

Preferably the hemicellulose is one which contains galactose, and the glycosyltransferase is membrane-bound *in vivo*.

The activity of the encoded polypeptide may be tested, for instance, by observing the addition of radiolabelled sugar residues from exogenously supplied radioactive sugar nucleotides to saccharides, for instance oligosaccharides, or more preferably polysaccharides. Such methods are described in more detail below.

Nucleic acid according to the present invention may include cDNA, RNA, genomic DNA and modified nucleic acids or nucleic acid analogs (e.g. peptide nucleic acid). Where a DNA sequence is specified, e.g. with reference to a figure, unless context requires otherwise the RNA equivalent, with U substituted for T where it occurs, is encompassed.

Nucleic acid molecules according to the present invention may be provided isolated and/or purified from their natural environment, in substantially pure or homogeneous form, or free or substantially free of other nucleic acids of the species of origin. Where used herein, the term "isolated" encompasses all of these possibilities.

The nucleic acid molecules may be wholly or partially synthetic. In particular they may be recombinant in that nucleic acid sequences which are not found together in nature (do not run contiguously) have been ligated or otherwise combined artificially. Alternatively they may have been synthesised directly e.g. using an automated synthesiser.

Most preferably the nucleic acid encodes a galactosyltransferase, which is capable of catalysing the biosynthesis of galactomannan.

Thus in one embodiment of this aspect of the invention, there is disclosed a nucleic acid comprising the nucleotide sequence shown in Seq ID No 1 (Annex 1a). This sequence represents that of a cDNA molecule encoding a galactosyl transferase gene from fenugreek. The encoded polypeptide (Seq ID No 2) is also shown in Annex 1b.

In a further embodiment of the invention, there is disclosed a nucleic

acid comprising the nucleotide sequence shown in Seq ID No 3 (Annex 2a). This sequence represents that of a cDNA molecule encoding a galactosyl transferase gene from guar. The encoded polypeptide (Seq ID No 4) is also shown in Annex 2b. These sequences are compared in Fig 1.

Also disclosed are nucleic acids which are variants of the sequences provided. A variant nucleic acid molecule shares homology with, or is identical to, all or part of the coding sequence discussed above.

Generally, variants may encode, or be used to isolate or amplify nucleic acids which encode, polypeptides which are capable of catalysing the biosynthesis of a complex non-cellulosic plant cell wall polysaccharide by binding nucleotide sugar precursors and transfer sugar residues to polysaccharides in the golgi compartment(s).

Such polypeptides may include not only galactosyltransferases, but also other (golgi located) glycosyltransferases e.g. those involved in galacto(gluco)mannan biosynthesis such as mannosyl and glucosyl transferases. Also included may be galactosyltransferases which act on pectin or xyloglucan.

Other polypeptides having the requisite characteristics may include arabinosyltransferase, glucosyltransferase, xylosyltransferase, mannosyltransferase, fucosyltransferase, rhamnosyltransferase, galacturonyltransferase and glucuronyltransferase.

Activities may conveniently be assessed using *in situ* analysis in chromatographic gels (e.g. agarose gels) containing a suitable substrate (e.g. galactomannan for galactosyltransferase activity). Such methods of assessment form one part of the present invention.

A typical method will comprise the steps of:
(i) applying a sample comprising a mixture of proteins to a detector gel, said detector gel comprising in admixture (a) a chromatographic gel suitable for chromatographic separation of a mixture of proteins; (b) an acceptor substrate for a glycosyltransferase, wherein the acceptor substrate is compatible with the chromatographic gel in that it does not impair the chromatographic properties of the gel, but is accessible as a substrate for the proteins of the mixture,
(ii) chromatographically separating said mixture on the basis of size and/or charge

(iii) locating the glycosyltransferase, if present, within the gel, on the basis of glycosyltransfer to the substrate.

Suitable 'compatible' substrates may include xyloglucan, xylan, glucomannan and pectin.

Variants of the present invention can include not only novel, naturally occurring, nucleic acids, isolatable using the sequences of the present invention, but also artificial nucleic acids having novel sequences, which can be prepared by the skilled person in the light of the present disclosure.

Thus a variant may be a distinctive part or fragment (however produced) corresponding to a portion of the sequence provided. The fragments may encode particular functional parts of the polypeptide, e.g. portions lacking the transmembrane α-helix near the N-terminus (e.g. between residues 15 to 41 of the fenugreek sequence, or as underlined in Fig 1) which may have improved properties such as solubility or activity.

Equally the fragments may have utility in probing for, or amplifying, the sequence provided or closely related ones. Suitable lengths of fragment, and conditions, for such processes are discussed in more detail below.

Also included are nucleic acids which have been extended at the 3' or 5' terminus. Also included are sequences e.g. genomic sequences, having additional, non-expressed, portions ('introns').

Sequence variants which occur naturally may include homologous galactosyltransferases from other species, alleles (which will include polymorphisms or mutations at one or more bases) or pseudoalleles (which may occur at closely linked loci to the galactosyl transferase gene from fenugreek). Also included within the scope of the present invention would be isogenes, or other homologous genes which may belong to the same family as the galactosyltransferase gene (e.g. galactoglucomannan synthases). Although these may occur at different genomic loci to the gene, they are likely to share conserved regions with it.

Artificial variants (derivatives) may be prepared by those skilled in the art, for instance by site directed or random mutagenesis, or by direct synthesis. Preferably the variant nucleic acid is generated

either directly or indirectly (e.g. via one or more amplification or replication steps) from an original nucleic acid having all or part of the sequence shown in Seq ID No 1 or 3. Preferably it encodes a polypeptide which is capable of catalysing the biosynthesis of a complex non-cellulosic plant cell wall polysaccharide.

The term 'variant' nucleic acid as used herein encompasses all of these possibilities. When used in the context of polypeptides or proteins it indicates the encoded expression product of the variant nucleic acid.

Some of the aspects of the present invention relating to variants will now be discussed in more detail.

Homology and activity

Similarity or homology may be as defined and determined by the TBLASTN program, of Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-10, which is in standard use in the art, or, and this may be preferred, the standard program BestFit, which is part of the Wisconsin Package, Version 8, September 1994, (Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA, Wisconsin 53711). BestFit makes an optimal alignment of the best segment of similarity between two sequences. Optimal alignments are found by inserting gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman.

Homology may be at the nucleotide sequence and/or encoded amino acid sequence level. Preferably, the nucleic acid and/or amino acid sequence shares at least about 50%, or 60%, or 70%, or 80% homology, most preferably at least about 90%, 95%, 96%, 97%, 98% or 99% homology.

Homology may be over the full-length of the relevant sequence shown herein, or may be over a part of it, preferably over a contiguous sequence of about or greater than about 20, 25, 30, 33, 40, 50, 67, 133, 167, 200, 233, 267, 300, 333, 400 or more amino acids or codons, compared with Seq ID Nos 1 to 4 as appropriate.

Thus a variant polypeptide in accordance with the present invention may include within the sequence shown in Seq ID No 2 or 4, a single amino acid or 2, 3, 4, 5, 6, 7, 8, or 9 changes, about 10, 15, 20, 30, 40 or 50 changes, or greater than about 50, 60, 70, 80 or 90 changes. In addition to one or more changes within the amino acid sequence shown, a variant polypeptide may include additional amino acids at the C-

terminus and/or N-terminus. Naturally, changes to the nucleic acid which make no difference to the encoded polypeptide (i.e. 'degeneratively equivalent') are included.

The activity of a variant polypeptide may be assessed by transformation into a host cell capable of expressing the nucleic acid of the invention. Methodology for such transformation is described in more detail below.

Production of derivatives

Thus in a further aspect of the invention there is disclosed a method of producing a derivative nucleic acid comprising the step of modifying the coding sequence of Seq ID No 1 or 3.

Changes to a sequence, to produce a derivative, may be by one or more of addition, insertion, deletion or substitution of one or more nucleotides in the nucleic acid, leading to the addition, insertion, deletion or substitution of one or more amino acids in the encoded polypeptide.

Changes may be desirable for a number of reasons, including introducing or removing the following features: restriction endonuclease sequences; codon usage; other sites which are required for post translation modification; cleavage sites in the encoded polypeptide; motifs in the encoded polypeptide for glycosylation, lipoylation etc. Leader or other targeting sequences (e.g. membrane or golgi locating sequences) may be added to the expressed protein to determine its location following expression. All of these may assist in efficiently cloning and expressing an active polypeptide in recombinant form (as described below).

Other desirable mutations may be random or site directed mutagenesis in order to alter the activity (e.g. specificity) or stability of the encoded polypeptide.

Changes may be by way of conservative variation, i.e. substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another, such as arginine for lysine, glutamic for aspartic acid, or glutamine for asparagine. As is well known to those skilled in the art, altering the primary structure of a polypeptide by a conservative substitution may not significantly alter the activity of that peptide

because the side-chain of the amino acid which is inserted into the sequence may be able to form similar bonds and contacts as the side chain of the amino acid which has been substituted out. This is so even when the substitution is in a region which is critical in determining the peptide's conformation. Figures 6A and 6B show the predicted secondary structures of the fenugreek polypeptide.

Also included are variants having non-conservative substitutions. As is well known to those skilled in the art, substitutions to regions of a peptide which are not critical in determining its conformation may not greatly affect its activity because they do not greatly alter the peptide's three dimensional structure. In regions which are critical in determining the peptides conformation or activity such changes may confer advantageous properties on the polypeptide. Indeed, changes such as those described above may confer slightly advantageous properties on the peptide e.g. altered stability or specificity.

Identification of variants

In a further aspect of the present invention there is provided a method of identifying and/or cloning a nucleic acid variant from a plant which method employs Seq ID No 1 or 3 or a derivative thereof.

In each case, if need be, clones or fragments identified in the search can be extended. For instance if it is suspected that they are incomplete, the original DNA source (e.g. a clone library, mRNA preparation etc.) can be revisited to isolate missing portions e.g. using sequences, probes or primers based on that portion which has already been obtained to identify other clones containing overlapping sequence.

In one embodiment, nucleotide sequence information provided herein may be used in a data-base (e.g. of expressed sequence tags, or sequence tagged sites) search to find homologous sequences, such as those which may become available in due course, and expression products of which can be tested for activity as described below.

In a further embodiment, a variant in accordance with the present invention is also obtainable by means of a method which includes:
(a) providing a preparation of nucleic acid, e.g. from plant cells,
(b) providing a nucleic acid molecule having a nucleotide sequence shown in or complementary to Seq ID No 1 or 3 or a derivative thereof,

(c) contacting nucleic acid in said preparation with said nucleic acid molecule under conditions for hybridisation of said nucleic acid molecule to any said gene or homologue in said preparation, and identifying said gene or homologue if present by its hybridisation with said nucleic acid molecule.

Probing may employ the standard Southern blotting technique. For instance DNA may be extracted from cells and digested with different restriction enzymes. Restriction fragments may then be separated by electrophoresis on an agarose gel, before denaturation and transfer to a nitrocellulose filter. Labelled probe may be hybridised to the DNA fragments on the filter and binding determined. DNA for probing may be prepared from RNA preparations from cells.

Test nucleic acid may be provided from a cell as genomic DNA, cDNA or RNA, or a mixture of any of these, preferably as a library in a suitable vector. If genomic DNA is used the probe may be used to identify untranscribed regions of the gene (e.g. promoters etc.), such as is described hereinafter. Probing may optionally be done by means of so-called 'nucleic acid chips' (see Marshall & Hodgson (1998) Nature Biotechnology 16: 27-31, for a review).

When using genomic DNA, this method may be used to isolate promoters or other regulatory elements, the activity of which may be confirmed by analogy with the methods below e.g. using promoterless constructs in which isolated fragments are operably linked to detectable genes.

Preliminary experiments may be performed by hybridising under low stringency conditions. For probing, preferred conditions are those which are stringent enough for there to be a simple pattern with a small number of hybridisations identified as positive which can be investigated further.

For instance, screening may initially be carried out under conditions, which comprise a temperature of about 37°C or less, a formamide concentration of less than about 50%, and a moderate to low salt (e.g. Standard Saline Citrate ('SSC') = 0.15 M sodium chloride; 0.15 M sodium citrate; pH 7) concentration. Alternatively, a temperature of about 50°C or less and a high salt (e.g. 'SSPE'= 0.180 mM sodium chloride; 9 mM disodium hydrogen phosphate; 9 mM sodium dihydrogen phosphate; 1 mM sodium EDTA; pH 7.4). Preferably the screening is carried out at about 37°C, a formamide concentration of about 20%, and a salt concentration

of about 5 X SSC, or a temperature of about 50°C and a salt concentration of about 2 X SSPE. These conditions will allow the identification of sequences which have a substantial degree of homology (similarity) with the probe sequence, without requiring the perfect homology for the identification of a stable hybrid.

Preferably, hybridisation conditions will be selected (e.g using higher temperatures) which allow the identification of sequences having 70% or more (e.g. 80%, 90%, 95%, 96%, 97%, 98% or 99%) sequence identity with the probe, while discriminating against sequences which have a lower level of sequence identity with respect to the probe. After low stringency hybridisation has been used to identify several nucleic acids having a substantial degree of similarity with the probe sequence, this subset is then subjected to high stringency hybridisation, so as to identify those clones having a particularly high level of homology with respect to the probe sequences. High stringency conditions comprise a temperature of about 42°C or less, a formamide concentration of less than about 20%, and a low salt (SSC) concentration. Alternatively they may comprise a temperature of about 65°C or less, and a low salt (SSPE) concentration. Preferred conditions for such screening comprise a temperature of about 42°C, a formamide concentration of about 20%, and a salt concentration of about 2 X SSC, or a temperature of about 65°C, and a salt concentration of about 0.2 SSPE.

It is well known in the art to increase stringency of hybridisation gradually until only a few positive clones remain. Suitable conditions would be achieved when a large number of hybridising fragments were obtained while the background hybridisation was low. Using these conditions nucleic acid libraries, e.g. cDNA libraries representative of expressed sequences, may be searched. Those skilled in the art are well able to employ suitable conditions of the desired stringency for selective hybridisation, taking into account factors such as oligonucleotide length and base composition, temperature and so on.

Binding of a probe to target nucleic acid (e.g. DNA) may be measured using any of a variety of techniques at the disposal of those skilled in the art. For instance, probes may be radioactively, fluorescently or enzymatically labelled. Other methods not employing labelling of probe include amplification using PCR (see below), RN'ase cleavage and allele specific oligonucleotide probing. The identification of successful hybridisation is followed by isolation of the nucleic acid

which has hybridised, which may involve one or more steps of PCR or amplification of a vector in a suitable host.

Amplification of variants

In a further embodiment, hybridisation of nucleic acid molecule to a variant may be determined or identified indirectly, e.g. using a nucleic acid amplification reaction, particularly the polymerase chain reaction (PCR). PCR requires the use of two primers to specifically amplify target nucleic acid, so preferably two nucleic acid molecules with sequences characteristic of glycosyltransferases are employed. Using RACE PCR, only one such primer may be needed (see "PCR protocols; A Guide to Methods and Applications", Eds. Innis et al, Academic Press, New York, (1990)).

Thus a method involving use of PCR in obtaining nucleic acid according to the present invention may include:

(a) providing a preparation of plant nucleic acid, e.g. from a seed or other appropriate tissue or organ,

(b) providing a pair of nucleic acid molecule primers useful in (i.e. suitable for) PCR, at least one said primer having a sequence shown in or complementary to a sequence shown in Seq ID No 1 or 3 or a derivative thereof,

(c) contacting nucleic acid in said preparation with said primers under conditions for performance of PCR,

(d) performing PCR and determining the presence or absence of an amplified PCR product. The presence of an amplified PCR product may indicate identification of a variant.

Nucleic acids for probing or amplification

An oligonucleotide for use in probing or PCR may be about 30 or fewer nucleotides in length (e.g. 18, 21 or 24). Generally specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16-24 nucleotides in length may be preferred. Those skilled in the art are well versed in the design of primers for use in processes such as PCR. If required, probing can be done with entire restriction fragments of the gene disclosed herein which may be 100's or even 1000's of nucleotides in length.

It may be desirable to select primers or probes which are distinctive for particular parts of the sequence which are likely to be associated with particular activities e.g. it may be desirable to avoid using sequence from the helix region as these are more likely to cross react

with sequences not forming part of the present invention.

As used hereinafter, unless the context demands otherwise, where "galactosyltransferase" is specified, the invention also covers corresponding applications employing the variants discussed above.

In one aspect of the present invention, the nucleic acid encoding the galactosyltransferase described above is in the form of a recombinant and preferably replicable vector.

"Vector" is defined to include, inter alia, any plasmid, cosmid, phage or *Agrobacterium* binary vector in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g. autonomous replicating plasmid with an origin of replication).

Specifically included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eucaryotic (e.g. higher plant, mammalian, yeast or fungal) cells.

A vector including nucleic acid according to the present invention need not include a promoter or other regulatory sequence, particularly if the vector is to be used to introduce the nucleic acid into cells for recombination into the genome.

Preferably the nucleic acid in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell such as a microbial, e.g. bacterial, or plant cell. The vector may be a bi-functional expression vector which functions in multiple hosts. In the case of genomic DNA, this may contain its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell

By "promoter" is meant a sequence of nucleotides from which transcription may be initiated of DNA operably linked downstream (i.e. in the 3' direction on the sense strand of double-stranded DNA).

"Operably linked" means joined as part of the same nucleic acid molecule, suitably positioned and oriented for transcription to be initiated from the promoter. DNA operably linked to a promoter is "under transcriptional initiation regulation" of the promoter.

Thus this aspect of the invention provides a gene construct, preferably a replicable vector, comprising a promoter operatively linked to a nucleotide sequence provided by the present invention, such as the fenugreek galactosyltransferase gene or a variant thereof.

Generally speaking, those skilled in the art are well able to construct vectors and design protocols for recombinant gene expression. Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. For further details see, for example, *Molecular Cloning: a Laboratory Manual:* 2nd edition, Sambrook *et al*, 1989, Cold Spring Harbor Laboratory Press.

Many known techniques and protocols for manipulation of nucleic acid, for example in preparation of nucleic acid constructs, mutagenesis (see above discussion in respect of variants), sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in Molecular Biology*, Second Edition, Ausubel et al. eds., John Wiley & Sons, 1992. The disclosures of Sambrook et al. and Ausubel et al. are incorporated herein by reference.

One embodiment of this aspect of the present invention provides a gene construct, preferably a replicable vector, comprising an inducible promoter operatively linked to a nucleotide sequence provided by the present invention, such as Seq ID No 1 or 3.

The term "inducible" as applied to a promoter is well understood by those skilled in the art. In essence, expression under the control of an inducible promoter is "switched on" or increased in response to an applied stimulus. The nature of the stimulus varies between promoters. Some inducible promoters cause little or undetectable levels of expression (or no expression) in the absence of the appropriate stimulus. Other inducible promoters cause detectable constitutive expression in the absence of the stimulus. Whatever the level of expression is in the absence of the stimulus, expression from any

inducible promoter is increased in the presence of the correct stimulus.

Particularly of interest in the present context are nucleic acid constructs which operate as plant vectors.

Specific procedures and vectors previously used with wide success upon plants are described by Guerineau and Mullineaux (1993) (Plant transformation and expression vectors. In: Plant Molecular Biology Labfax (Croy RRD ed) Oxford, BIOS Scientific Publishers, pp 121-148).

Suitable promoters which operate in plants include the Cauliflower Mosaic Virus 35S (CaMV 35S); the cauliflower meri 5 and the *Arabidopsis thaliana* LEAFY promoter that is expressed very early in flower development. Other promoters include the rice actin promoter. Inducible promoters may include the GST-II-27 gene promoter which has been shown to be induced by certain chemical compounds which can be applied to growing plants. The promoter is functional in both monocotyledons and dicotyledons. Other examples are disclosed at pg 120 of Lindsey & Jones (1989) "Plant Biotechnology in Agriculture" Pub. OU Press, Milton Keynes, UK. The promoter may be selected to include one or more sequence motifs or elements conferring developmental and/or tissue-specific regulatory control of expression.

If desired, selectable genetic markers may be included in the construct, such as those that confer selectable phenotypes such as resistance to antibiotics or herbicides (e.g. kanamycin, hygromycin, phosphinotricin, chlorsulfuron, methotrexate, gentamycin, spectinomycin, imidazolinones and glyphosate).

The present invention also provides methods comprising introduction of such a construct into a cell and/or induction of expression of a construct within a cell, by application of a suitable stimulus e.g. an effective exogenous inducer.

In a further aspect of the invention, there is disclosed a host cell containing a heterologous construct according to the present invention, especially a plant or a microbial cell (e.g. yeast cell).

The term "heterologous" is used broadly in this aspect to indicate that the gene/sequence of nucleotides in question (e.g. encoding galactosyltransferase) have been introduced into said cells of the

plant or an ancestor thereof, using genetic engineering, i.e. by human intervention. A heterologous gene may replace an endogenous equivalent gene, i.e. one which normally performs the same or a similar function, or the inserted sequence may be additional to the endogenous gene or other sequence. Nucleic acid heterologous to a plant cell may be non-naturally occurring in cells of that type, variety or species. Thus the heterologous nucleic acid may comprise a coding sequence of or derived from a particular type of plant cell or species or variety of plant, placed within the context of a plant cell of a different type or species or variety of plant. A further possibility is for a nucleic acid sequence to be placed within a cell in which it or a homolog is found naturally, but wherein the nucleic acid sequence is linked and/or adjacent to nucleic acid which does not occur naturally within the cell, or cells of that type or species or variety of plant, such as operably linked to one or more regulatory sequences, such as a promoter sequence, for control of expression.

The host cell (e.g. plant cell) is preferably transformed by the construct, which is to say that the construct becomes established within the cell, altering one or more of the cell's characteristics and hence phenotype e.g. with respect to CCWP production.

Nucleic acid can be transformed into plant cells using any suitable technology, such as a disarmed Ti-plasmid vector carried by Agrobacterium exploiting its natural gene transfer ability (EP-A-270355, EP-A-0116718, NAR 12(22) 8711 - 87215 1984), particle or microprojectile bombardment (US 5100792, EP-A-444882, EP-A-434616) microinjection (WO 92/09696, WO 94/00583, EP 331083, EP 175966, Green *et al.* (1987) *Plant Tissue and Cell Culture*, Academic Press), electroporation (EP 290395, WO 8706614 Gelvin Debeyser) other forms of direct DNA uptake (DE 4005152, WO 9012096, US 4684611), liposome mediated DNA uptake (e.g. Freeman *et al. Plant Cell Physiol.* 29: 1353 (1984)), or the vortexing method (e.g. Kindle, *PNAS U.S.A.* 87: 1228 (1990d) Physical methods for the transformation of plant cells are reviewed in Oard, 1991, *Biotech. Adv.* 9: 1-11.

Agrobacterium transformation is widely used by those skilled in the art to transform dicotyledonous species. Recently, there has been substantial progress towards the routine production of stable, fertile transgenic plants in almost all economically relevant monocot plants (see e.g. Hiei *et al.* (1994) *The Plant Journal* 6, 271-282)). Microprojectile bombardment, electroporation and direct DNA uptake are

preferred where Agrobacterium alone is inefficient or ineffective. Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, eg bombardment with Agrobacterium coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with Agrobacterium (EP-A-486233).

The particular choice of a transformation technology will be determined by its efficiency to transform certain plant species as well as the experience and preference of the person practising the invention with a particular methodology of choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce nucleic acid into plant cells is not essential to or a limitation of the invention, nor is the choice of technique for plant regeneration.

Thus a further aspect of the present invention provides a method of transforming a plant cell involving introduction of a construct as described above into a plant cell and causing or allowing recombination between the vector and the plant cell genome to introduce a nucleic acid according to the present invention into the genome.

The invention further encompasses a host cell transformed with nucleic acid or a vector according to the present invention (e.g comprising the galactosyltransferase sequence) especially a plant or a microbial cell. In the transgenic plant cell (i.e. transgenic for the nucleic acid in question) the transgene may be on an extra-genomic vector or incorporated, preferably stably, into the genome. There may be more than one heterologous nucleotide sequence per haploid genome.

Generally speaking, following transformation, a plant may be regenerated, e.g. from single cells, callus tissue or leaf discs, as is standard in the art. Almost any plant can be entirely regenerated from cells, tissues and organs of the plant. Available techniques are reviewed in Vasil et al., *Cell Culture and Somatic Cell Genetics of Plants, Vol I, II and III, Laboratory Procedures and Their Applications*, Academic Press, 1984, and Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989.

The generation of fertile transgenic plants has been achieved in the cereals rice, maize, wheat, oat, and barley (reviewed in Shimamoto, K. (1994) *Current Opinion in Biotechnology* 5, 158-162.; Vasil, *et al.*

(1992) *Bio/Technology* 10, 667-674; Vain et al., 1995, *Biotechnology Advances* 13 (4): 653-671; Vasil, 1996, *Nature Biotechnology* 14 page 702).

Plants which include a plant cell according to the invention are also provided.

A plant according to the present invention may be one which does not breed true in one or more properties. Plant varieties may be excluded, particularly registrable plant varieties according to Plant Breeders' Rights. It is noted that a plant need not be considered a "plant variety" simply because it contains stably within its genome a transgene, introduced into a cell of the plant or an ancestor thereof.

In addition to the regenerated plant, the present invention embraces all of the following: a clone of such a plant, seed, selfed or hybrid progeny and descendants (e.g. F1 and F2 descendants) and any part of any of these. The invention also provides a plant propagule from such a plant, that is any part which may be used in reproduction or propagation, sexual or asexual, including cuttings, seed and so on.

Preferably the plant is an endospermic legume which contains galactomannan as a CWSP. One example is the guar plant. Some methods for transforming and regenerating such plants are discussed in WO97/20937 (Danisco).

The present invention also encompasses the expression product of any of the galactosyltransferase or variant nucleic acid sequences disclosed above, and also methods of making the expression product by expression from encoding nucleic acid therefore under suitable conditions, which may be in suitable host cells.

Particularly included is a truncated polypeptide, lacking the transmembrane helix, which is soluble and not membrane-associated and which also has galactosyltransferase activity.

Following expression, the product may be isolated from the expression system (e.g. microbial) and may be used as desired, for instance in formulation of a composition including at least one additional component.

Alternatively the product may be used to perform its function *in vivo*

and in particular *in planta* as discussed above.

Purified galactosyltransferase protein, or a variant thereof, may be used to raise antibodies employing techniques which are standard in the art. Antibodies and polypeptides comprising antigen-binding fragments of antibodies may be used in identifying variants from other species as discussed further below.

Methods of producing antibodies include immunising a mammal (e.g. human, mouse, rat, rabbit, horse, goat, sheep or monkey) with the protein or a fragment thereof. Antibodies may be obtained from immunised animals using any of a variety of techniques known in the art, and might be screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or immunoprecipitation may be used (Armitage et al, 1992, Nature 357: 80-82). Antibodies may be polyclonal or monoclonal. Single chain antibodies e.g. from *Camelidae* may be preferred (see WO 94/25591 of Unilever).

Antibodies may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any specific binding substance having a binding domain with the required specificity. Thus, this term covers antibody fragments, derivatives, functional equivalents and homologues of antibodies, including any polypeptide comprising an immunoglobulin binding domain, whether natural or synthetic. Chimaeric molecules comprising an immunoglobulin binding domain, or equivalent, fused to another polypeptide are therefore included. Cloning and expression of Chimaeric antibodies are described in EP-A-0120694 and EP-A-0125023. It has been shown that fragments of a whole antibody can perform the function of binding antigens. Examples of binding fragments are (I) the Fab fragment consisting of VL, VH, CL and CH1 domains; (ii) the Fd fragment consisting of the VH and CH1 domains; (iii) the Fv fragment consisting of the Vl and VH domains of a single antibody; (iv) the dAb fragment (Ward, E.S. et al., Nature 341, 544-546 (1989) which consists of a VH domain; (v) isolated CDR regions; (vi) F(ab')2 fragments, a bivalent fragment comprising two linked Fab fragments (vii) single chain Fv molecules (scFv), wherein a VH domain and a VL domain are linked by a peptide linker which allows the two domains to associate to form an antigen binding site (Bird et al, Science, 242, 423-426, 1988; Huston et al, PNAS USA, 85, 5879-5883, 1988); (viii) bispecific single chain Fv dimers (PCT/US92/09965) and (ix) "diabodies", multivalent or multispecific fragments constructed by

gene fusion (WO94/13804; P Holliger et al Proc. Natl. Acad. Sci. USA 90 6444-6448, 1993).

Diabodies are multimers of polypeptides, each polypeptide comprising a first domain comprising a binding region of an immunoglobulin light chain and a second domain comprising a binding region of an immunoglobulin heavy chain, the two domains being linked (e.g. by a peptide linker) but unable to associate with each other to form an antigen binding site: antigen binding sites are formed by the association of the first domain of one polypeptide within the multimer with the second domain of another polypeptide within the multimer (WO94/13804).

As an alternative or supplement to immunising a mammal, antibodies with appropriate binding specificity may be obtained from a recombinantly produced library of expressed immunoglobulin variable domains, e.g. using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047.

Antibodies raised to a polypeptide or peptide can be used in the identification and/or isolation of variant polypeptides, and then their encoding genes. Thus, the present invention provides a method of identifying or isolating a galactosyltransferase or variant thereof (as discussed above), comprising screening candidate polypeptides with a polypeptide comprising the antigen-binding domain of an antibody (for example whole antibody or a fragment thereof) which is able to bind said galactosyltransferase polypeptide or variant thereof, or preferably has binding specificity for such a polypeptide. Specific binding members such as antibodies and polypeptides comprising antigen binding domains of antibodies that bind and are preferably specific for a galactosyltransferase polypeptide or mutant or derivative thereof represent further aspects of the present invention, as do their use and methods which employ them.

Candidate polypeptides for screening may for instance be the products of an expression library created using nucleic acid derived from an plant of interest, or may be the product of a purification process from a natural source. A polypeptide found to bind the antibody may be isolated and then may be subject to amino acid sequencing. Any suitable technique may be used to sequence the polypeptide either wholly or partially (for instance a fragment of the polypeptide may be

sequenced). Amino acid sequence information may be used in obtaining nucleic acid encoding the polypeptide, for instance by designing one or more oligonucleotides (e.g. a degenerate pool of oligonucleotides) for use as probes or primers in hybridization to candidate nucleic acid.

In addition to the aspects above, the invention further provides use of the materials described herein for altering the quality and/or quantity of CWSP in a host cell. Particularly for altering the mannose:galactose ratio in a mannose/galactose containing compound in that host cell.

For instance it provides a method of influencing or affecting the CWSP content of a host cell (preferably a plant cell), comprising the step of causing or allowing expression of a heterologous nucleic acid sequence encoding a biosynthetic enzyme as discussed above within the cell.

In addition to the aspects above, the invention further provides a method of influencing or affecting the glycosyltransferase activity in a plant, the method comprising the step of causing or allowing expression of a heterologous nucleic acid sequence as discussed above (e.g. encoding the fenugreek or guar galactosyltransferase or a variant thereof) within the cells of the plant.

In each case the step may be preceded by the earlier step of introduction of the nucleic acid into a cell of the plant or an ancestor thereof.

The foregoing discussion has been generally concerned with uses of the nucleic acids of the present invention for production of functional polypeptides, for instance for the purpose of increasing the galactosyltransferase activity in the cell.

However the information disclosed herein may also be used to reduce the activity of galactosyltransferases in cells in which it is desired to do so.

For instance down-regulation of expression of a target gene may be achieved using anti-sense technology.

In using anti-sense genes or partial gene sequences to down-regulate gene expression, a nucleotide sequence is placed under the control of a

promoter in a "reverse orientation" such that transcription yields RNA which is complementary to normal mRNA transcribed from the "sense" strand of the target gene. See, for example, Rothstein et al, 1987; Smith *et al*,(1988) *Nature* 334, 724-726; Zhang *et al*,(1992) *The Plant Cell* 4, 1575-1588, English *et al*., (1996) *The Plant Cell* 8, 179-188. Antisense technology is also reviewed in Bourque, (1995), *Plant Science* 105, 125-149, and Flavell, (1994) *PNAS USA* 91, 3490-3496.

Thus a nucleotide sequence which is complementary to any of the coding sequences discussed above (including variants) forms one part of the present invention.

"Complementary to" means capable of base pairing with, whereby A is the complement of T (and U); G is the complement of C.

An alternative to anti-sense is to use a copy of all or part of the gene (galactosyltransferase or variant) inserted in sense, that is the same, orientation as the natural gene, to achieve reduction in expression of the target gene by co-suppression. See, for example, van der Krol *et al*., (1990) *The Plant Cell* 2, 291-299; Napoli *et al*., (1990) *The Plant Cell* 2, 279-289; Zhang *et al*., (1992) *The Plant Cell* 4, 1575-1588, and US-A-5,231,020. Further refinements of the gene silencing or co-suppression technology may be found in WO95/34668 (Biosource); Angell & Baulcombe (1997) The EMBO Journal 16,12:3675-3684; and Voinnet & Baulcombe (1997) Nature 389: pg 553.

Further options for down regulation of gene expression include the use of ribozymes, e.g. hammerhead ribozymes, which can catalyse the site-specific cleavage of RNA, such as mRNA (see e.g. Jaeger (1997) "The new world of ribozymes" Curr Opin Struct Biol 7:324-335, or Gibson & Shillitoe (1997) "Ribozymes: their functions and strategies for their use" Mol Biotechnol 7: 242-251.)

The complete sequence corresponding to the coding sequence (in reverse orientation for anti-sense) need not be used. For example fragments of sufficient length may be used. It is a routine matter for the person skilled in the art to screen fragments of various sizes and from various parts of the coding sequence to optimise the level of anti-sense inhibition. It may be advantageous to include the initiating methionine ATG codon, and perhaps one or more nucleotides upstream of the initiating codon. A further possibility is to target a conserved

sequence of a gene, e.g. a sequence that is characteristic of one or more genes, such as a regulatory sequence.

The sequence employed may be about 500 nucleotides or less, possibly about 400 nucleotides, about 300 nucleotides, about 200 nucleotides, or about 100 nucleotides. It may be possible to use oligonucleotides of much shorter lengths, 14-23 nucleotides, although longer fragments, and generally even longer than about 500 nucleotides are preferable where possible, such as longer than about 600 nucleotides, than about 700 nucleotides, than about 800 nucleotides, than about 1000 nucleotides or more.

It may be preferable that there is complete sequence identity in the sequence used for down-regulation of expression of a target sequence, and the target sequence, although total complementarity or similarity of sequence is not essential. One or more nucleotides may differ in the sequence used from the target gene. Thus, a sequence employed in a down-regulation of gene expression in accordance with the present invention may be a wild-type sequence (e.g. gene) selected from those available, or a variant of such a sequence.

The sequence need not include an open reading frame or specify an RNA that would be translatable. It may be preferred for there to be sufficient homology for the respective anti-sense and sense RNA molecules to hybridise. There may be down regulation of gene expression even where there is about 5%, 10%, 15% or 20% or more mismatch between the sequence used and the target gene. Effectively, the homology should be sufficient for the down-regulation of gene expression to take place.

Thus the present invention further provides the use of Seq ID No 1 or 3, or the complement thereof, or a variant of any of these, for down-regulation of gene expression, particularly down-regulation of expression of a galactosyltransferase gene or variant thereof, preferably in order to influence the galactosyltransferase activity in a host cell, more preferably a plant cell, most preferably a plant.

The invention further provides use of an antibody to achieve the same.

Anti-sense or sense regulation may itself be regulated by employing an inducible promoter in an appropriate construct.

A yet further method of manipulating galactosyltransferase activity is to express an antibody to the enzyme in the plant. It has been demonstrated that functional antibodies and antibody fragments can be expressed intracellularly and can be targeted to sub-cellular compartments. Alteration of phenotype by this method has been demonstrated, for instance by Artsaenko et al (1995) Plant J 8: 745-750 and Owen et al (1992) Bio/Technology 10: 790-794.

In a further aspect of the present invention there is disclosed a plant product derived from any of the transformed plants or plant cells, or produced by any of the methods, discussed above in relation to other aspects of the invention (e.g. in which galactosyltransferase activity has been altered).

Preferably the plant product comprises an altered galactomannan, which is to say that the galactomannan contains an altered (preferably reduced) ratio of galactose to mannose and/or an altered backbone galactose distribution.

In a further aspect of the present invention there is provided a commodity comprising the plant product described above (e.g. up to 5%, preferably 0.1 - 3%), particularly a human or animal foodstuff, or a cosmetic.

Particularly envisaged in terms of human foodstuffs is a frozen food product, for instance an ice cream or water ice. Also of interest are salad dressings, sauces, gelled desserts and "reduced-fat" products.

Animal foodstuffs may include gel-based petfoods.

The food composition comprising altered galactomannan plus one other polysaccharide selected from: xanthan; carrageenan; agarose.

Galactomannans having altered hydrophilic and cryogelation properties may have particular application to industry as additives e.g. as stabilisers, emulsifiers, and in combination with other polysaccharides, to impart more complex rheologies.

The various aspects of the invention will now be further described with reference to the following non-limiting Figures and Examples. Other embodiments falling within the scope of the present invention will occur to those skilled in the art in the light of these.

FIGURES

Figure 1: shows an alignment of the fenugreek galactosyltransferase with the putative guar galactosyltransferase sequence. Galtran2.pro = fenugreek galactosyltransferase; Guargalt.PRO = guar sequence. Residues matching the fenugreek sequence exactly are boxed. Numbering corresponds to the guar sequence. Predicted membrane-spanning a-helix is underlined.

Figure 2: typical data-sets correlating galactosyltransferase activity and a 51K protein. Triton X-100 solubilised extracts were separated on IEF gels containing the same detergent. Strips from each gel were treated to localise galactosyl-transferase activity and separated protein, and to plot the pH gradient.

A and B. Alignment of galactosyltransferase profile activity with second-dimension SDS-PAGE. Activity correlates closely with a 51K protein [50K position arrowed].

C. SDS-PAGE separation [welled gel] of slices from a further strip of the same IEF gel [50K position arrowed]. The two peak activity slices [see A], indicated with asterisks, are enriched in a 51K protein.

D and E. Second dimension SDS-PAGE and Western blot of an identical gel challenged with an antiserum raised against pea vicilin. The position of the 51K protein is arrowed in each.

Figure 3: cDNA and deduced protein sequence of c.500 bp clone obtained by 3'RACE. The sequences of the degenerate gene-specific primer and an antisense primer [GTPA3] are double underlined. Known sequences from the 51 K protein are underlined and italicised.

Figure 4: cDNA and deduced protein sequence from c.1000 bp clone. The upper sequence is from the 5' end of the clone, the lower from the 3' end. The sequences of the 5' and 3' degenerate primers used to amplify the cDNA are double underlined. The known protein sequence from the 51 K protein is underlined and italicised.

Figure 5: cDNA and deduced protein sequence from c.1500 bp clone. The primers used to amplify the cDNA are double underlined. The known protein sequence from the 51 K protein is underlined. The orf beginning

at bp1 and ending at bp 1314 encodes a 438 aa protein.

Figure 6: Secondary structure prediction using neural network based program. 6A shows predicted helix (H); extended sheet (E); other loop (blank). 6B shows predicted transmembrane helix (T).

Figure 7: Digital autoradiogram of *endo*-β-D-mannanase digests of labelled polymeric products formed during galactosyltransferase assays of 10x concentrated supernatants from three different *Pichia* colonies carrying truncated constructs. A trace amount of the galactomannan active α-galactosidase from guar seeds was added to the digest in the centre lane. Gal = Galactose. Abbreviations for diagnostic galactomanno-oligosaccharides (Reid et al, 1995): M2G = galactosylmannobiose; M3G = galactosylmannotriose; M5G2 = digalactosylmannopentaose; O = galactomanno-octasaccharides; N= galactomanno-nonasaccharides.

SEQUENCE ANNEXES

Annexe 1a fenugreek cDNA sequence - Seq ID No 1

Annex 1b: translation of the fenugreek cDNA sequence - Seq ID No 2

Annexe 2a: guar cDNA sequence - Seq ID No 3

Annex 2b: translation of the guar cDNA sequence - Seq ID No 4

EXAMPLES

Example 1- Identification of a polypeptide and acquisition of amino acid sequence

Isolation of membranes capable of catalysing galactomannan biosynthesis in vitro from developing fenugreek seeds.

Fenugreek plants were grown to flowering and fruiting under conditions which have been described elsewhere (Edwards et al. 1992). Membranes were prepared using a method similar to that described previously (Edwards et al. 1989). Endosperms were hand-isolated at a stage of seed development during which intensive galactomannan biosynthesis was taking place [35-40 days after anthesis under our growth conditions],

and homogenised in a glass Potter homogeniser with 50 mM Tris HCl buffer pH 7.5 containing 1 mM EDTA and 5 mM dithiothreitol [DTT]. Usually the ratio of buffer to plant tissue was 0.5 ml buffer per endosperm. Larger particles were removed by an initial centrifugation at 13000 g [10 min], and the supernatant was spun at 100000 g for 1 hour. The pelleted membranes were resuspended in the same buffer [usually 100 μl per endosperm].

Standard assays for mannan synthase and galactosyltransferase in isolated membranes

These have been described (Edwards et al, 1989, 1992, 1995). Briefly, the incubation mixture [total volume 100 μl; incubated at 30°C, usually for 1 h] comprised membranes [usually equivalent to 0.1 - 1 endosperm], DTT [2.5 mM], EDTA [0.5 mM], MgCl2 [2.5 mM], MnCl2 [5 mM], UDP-Gal [800 μM] and/or GDP-Man [80 μM] in 25 mM Tris HCl buffer, pH 7.5. The GDP-Man and/or the UDP-Gal substrate was labelled with the appropriate nucleoside diphospho-[U-^{14}C]-sugars. Specific radioactivities were adjusted to 25-250 Bq.nmol^{-1} and were checked by scintillation counting in each experiment. At the end of the incubation time glacial acetic acid [50 μl] was added and the mixture heated at 100°C for 2 min. Carrier galactomannan [100 μl of a 0.2 % w/v solution of locust bean galactomannan] was then added, followed by methanol to a final concentration of 70 % v/v. The mixture was heated [70°C for 10 min] and centrifuged [13000 g, 10 min]. the supernatant was discarded, and the pellet washed twice with hot 70% methanol as has been described (Edwards et al. 1989).

Mannan synthase could be assayed as above using labelled GDP-Man and unlabelled UDP-Gal. Under these conditions the product was a galactomannan, labelled in the mannosyl residues. It could be assayed also in the absence of unlabelled UDP-Gal, when the product was labelled (1-4)-β-mannan.

Galactosyltransferase was assayed using labelled UDP-Gal and unlabelled GDP-Man. It could not be assayed in the absence of GDP-Man, since the galactosyl residues were transferred only to newly transferred mannose residues (Edwards et al. 1989, 1992, 1995)

Detergent treatment of the membranes

Membranes were isolated as above, and resuspended [homogeniser] in 100

mM Tris HCl buffer pH 7.5 [12.5 (μl per endosperm] containing EDTA [2 mM] and DTT [10 mM]. Samples of the resuspended membranes were mixed with an equal volume of 2% [w/v] detergent, placed on ice and homogenised briefly every 10 min for 30 min. Suspensions were then centrifuged at 100000 g for 1 hour. Supernatants were retained, and pellets were resuspended in a 1:1 mixture of resuspension buffer and 2% detergent, with a volume equal to that of the supernatant. Standard assays for mannan synthase and galactosyltransferase were carried out on supernatants and resuspended pellets.

Of the detergents used [digitonin, CHAPS, octyl glucoside, Triton X-100, NP-40], only digitonin gave appreciable mannan synthase activity in the 100000 g supernatant.

In a typical digitonin experiment, the activity surviving digitonin treatment was 12.4%, of which 35% was in the supernatant and 65% in the pellet. None of the detergents gave any galactosyltransferase activity, either in pellets or supernatants.

Demonstration of galactomannan galactosyltransferase in digitonin extracts using manno-oligosaccharide acceptors.

Our observation that mannan synthase activity was retained in digitonin extracts without associated galactosyltransferase activity indicated either that the galactosyltransferase activity had been denatured preferentially or that the functional association between the mannan synthase and the galactosyltransferase had been disrupted by detergent treatment to the extent that the nascent mannan backbone was no longer available to the galactosyltransferase as acceptor substrate. If the latter were true it was possible that replacement acceptor substrates could be added to the extracts to mimic the mannan backbone. Initially digitonin extracts exhibiting soluble mannan synthase activity were incubated as for the standard assay, with added mannohexaose [1 mM], no GDP-Man, and labelled UDP-Gal [800 μM]. At the end of the incubation period the mixture was diluted by the addition of water [100 μl] and then spun through small columns (approximately 200 μl) of DEAE-cellulose [Whatman DE52] anion-exchanger which had been equilibrated with buffer identical to that used in the incubation. This procedure removed almost all of the unused labelled UDP-Gal substrate, which is negatively charged and binds to the cationic DEAE cellulose. After freeze-drying, the column eluate was dissolved in water (50 μl) and 20μl samples were spotted onto silicagel TLC plates (Merck 5553).

The plates were developed three times in a solvent composed of n-propanol, nitromethane and water (5:2:3 by vol), dried and analysed by digital autoradiography. The appearance of a radioactive spot running slightly slower than mannohexaose indicated that labelled galactose had been transferred from UDP-Gal to the mannohexaose. A pure sample of the labelled compound was obtained by carrying out a larger scale incubation and column purification as above, and strip-loading TLC plates with the column eluate. After developing the plates, the labelled product was located by digital autoradiography and then purified by removing the appropriate area of silica gel from the plates and eluting the silica gel with water. A pure α-galactosidase from guar seeds catalysed the complete conversion of the purified labelled compound to labelled galactose, and when the reaction was carried out in a graded fashion there were no labelled intermediates produced. Thus the labelled product carried a single α-linked galactose residue. Further analysis of the labelled product with a pure structure-sensitive endo-mannanase and a commercially available exo-β-mannosidase from snail [Sigma M9400] confirmed that the galactose residue had been transferred α-(1-6) to mannohexaose. The effectiveness of manno-oligosaccharides of different chain-length as galactosyl acceptors was compared [M5<<M6<M7<M8~M9, M10] and the nature of the products formed in each case was investigated using the three enzymes mentioned above, TLC and digital autoradiography. Results were consistent with a model for acceptor substrate binding, according to which the α-galactosyltransferase has an acceptor substrate binding requirement comprising six principal binding sites for mannosyl residues of the acceptor substrate. For transfer to occur, at least five of the sites must be occupied, and transfer occurs to the mannose residue at the third binding site [measured from the non-reducing end].

Thus manno-oligosaccharide acceptors allowed the assay of the galactomannan galactosyltransferase after digitonin solubilisation. Standard procedure was to incubate the detergent extract with mannohexaose [1 mM], MnCl2 [usually 10 mM] and labelled UDP-Gal [800 μM], dilute, spin through DEAE cellulose columns, freeze-dry the eluent and dissolve in water [50 μl] as above. Scintillation counting of an aliquot of the resulting solution gave a measure of the total radioactivity eluted from the column. The proportion of this activity present in the galactosylmanno-hexaose product of the galactosyltransferase reaction was estimated by TLC and quantitative digital autoradiography of a further aliquot.

Development of a new method to associate enzyme activity with particular proteins in the digitonin extract.

The small amounts of tissue available from the hand-dissected endosperm tissues, and the presence of the detergent in the extracts, made conventional procedures for protein purification or enrichment impracticable. A method to associate mannan synthase and galactosyltransferase activities with discrete polypeptides separated on SDS-PAGE gels was therefore devised. The method, with its extension and refinements as described below, was used successfully to pinpoint the galactomannan galactosyltransferase protein. Its effectiveness is strongly dependent upon its ability to give an exact correlation between enzyme activity profiles from IEF carried out in a first dimension, and SDS-PAGE carried out in the second dimension.

Generally, the method involved isoelectric focussing [IEF] of enzymatically active detergent extracts on vertical agarose minigels prepared in the presence of the solubilising detergent. It was found that detergent-solubilised proteins, present presumably in micelles, moved into the gel and were focussed according to their apparent pI values. Moreover, mannan synthase and galactosyltransferase activities in digitonin extracts were retained after focussing. After focussing, gels were cut into 1cm wide strips parallel to the direction of current flow. To determine the shape of the pH gradient, one such strip could be cut into slices perpendicular to the direction of current flow, each slice eluted with 1M KCl and the pH values of the resulting solutions measured. The pH gradient [establishment, shape, stability] was monitored also during focussing by loading the IEF gels with small samples of coloured "marker" proteins flanking the sample of detergent-solubilised enzyme. To measure galactosyltransferase activity on the IEF gels, further strips were cut into slices [usually 2mm] and each slice was assayed for activity.

In this way activity could be localised within the IEF gels. To determine which proteins were focussed at particular points within the gel, two related experimental approaches were used. In the first an IEF gel strip adjacent to the one sliced for activity determination was sliced in exactly the same way and each slice was treated with SDS-PAGE sample buffer and placed in an individual sample well of an SDS gel. Staining of the gel after SDS-PAGE then allowed a visual correlation of enzyme activity with polypeptide distribution. The second approach was to place an IEF strip adjacent to the one sliced for enzyme activity

determination along a long sample well of an SDS gel and subjecting it to SDS-PAGE in a direction perpendicular to that of IEF. This 2-dimensional IEF/SDS-PAGE approach gave an excellent visual correlation between proteins on the stained SDS-PAGE gel and enzyme activity.

In more detail; isoelectric focussing gels [8 x 10 cm] were prepared by assembling a "sandwich" of a glass plate to which pre-marked GelBond agarose gel support medium 0.1 mm thick [FMC BioProducts] had been attached [Hoefer Technical bulletin No 134], 1 mm spacers, and a notched alumina plate [Hoefer] in a Hoefer Gel Caster SE 245. GelBond was used to ensure that gel dimensions did not change during any manipulations and staining procedures. To aid subsequent division of gels, the reverse [hydrophobic, adjacent to the glass] side of the Gelbond film was pre-marked using a fine marker-pen with guide lines to facilitate accurate cutting into strips and slices. The agarose separation gel was prepared by mixing IsoGel agarose [120 mg], sorbitol [2.4 g] and water [10.36 ml] and heating on a boiling water bath for 10 min with frequent mixing to dissolve the agarose. After cooling to 65°C, the volume was made up to the original value. For digitonin gels 600 µl of 2% [w/v] digitonin [Sigma D1407] was added before boiling to give a final concentration of 0.1% in the gels. For Triton X-100 gels (see below), 600 µl of a 2% [w/v] solution of the detergent [Boehringer 789 704] was added after cooling to 65°C, due to the low [65°C] cloud point of this detergent, again giving a final detergent concentration in the gel of 0.1%. The in-gel detergent concentrations were above the critical micelle concentrations [CMC] of the detergents [0.09% for digitonin and 0.02% for Triton X-100] and were used to maintain protein solubility during IEF. Finally [at 65°C] 600 µl of ampholytes [a 4:1 (vol:vol) mixture of pH 5.0 - 8.0 Ampholine, Sigma A5799 and pH 3.5 - 10.0 Ampholine, Sigma A5174] were added to the agarose mixture to give a concentration of 2% in the gel. The gel "sandwich" was pre-warmed in an oven, the gel mixture [at 65°C] was added using a syringe, and a reference well comb [Hoefer] was inserted. This comb gives a 6.7 cm wide sample well, with a small 0.5 cm wide reference well alongside. The gel was left to set for 1 hour before it was assembled into a Hoefer SE 250 vertical gel apparatus which was cooled by water circulation to approximately 4°C. Cooling was also carried out during IEF to ensure adequate dissipation of heat generated and minimise loss of enzyme activity. The sample and reference wells were cleaned and dried using strips of filter paper, and the sample, overlay and IEF standards applied. The sample consisted of 750 µl of detergent extract

[detergent concentration 1%] prepared as described above, mixed with 45 µl of the same ampholyte mixture as was used to prepare the separating gel, 65 µl glycerol and 5 µl bromophenol blue [0.05 % w/v in water]. It was pipetted into the sample well of the gel. An overlay was prepared from 40 µl of the ampholyte mixture, 40 µl of 2% detergent, 40 µl glycerol 5 µl of the bromophenol blue solution and 680 µl water. A portion of this was pipetted into the reference well, and the remainder was layered carefully over the sample. The overlay is less dense than the sample, but more dense than the cathode buffer, thus forming a barrier to direct mixing of the sample and the strongly alkaline cathode buffer. Coloured IEF standards [Bio-Rad] (2.5 µl) were pipetted directly into the reference well. Finally, cathode buffer (20 mM NaOH) was carefully poured into the back (upper) chamber of the apparatus so that it did not mix with the overlay and sample, and anode buffer (6 mM phosphoric acid) poured into the lower chamber. The IEF was run at 200 V for 30 minutes during which most of the sample could be seen to enter the gel, and then at 600 V for 60 minutes. During this time the coloured IEF standards could be seen to migrate, focus and stabilise in position, and the current taken fell from about 12 mA to a stable final value of around 2 mA. After running, the gel sandwich was removed from the apparatus and the gel, attached to GelBond, separated from the plates and spacers. It was then cut up into strips parallel to the direction of current flow. The two extreme end strips were cut to include side-strips from the sample area. Thus one of them also included the reference standards. These two sections were fixed in 10% trichloroacetic acid [TCA] / 40% methanol for 15 min. During this time, two non-coloured standards at pI 6.0 and pI 6.5 became visible as opaque bands. This allowed them to be used in some experiments as markers for the peak of activity of galactosyltransferase [pI 6.0 in digitonin and pI 6.5 in Triton X-100]. The two strips were then dehydrated in methanol for 15 min, dried between sheets of filter paper and stained with Coomassie Blue. The stained strips showed the complete range of IEF standards. They also revealed the positions of stained bands in the sample, and showed whether or not the sample had focussed in bands running perfectly horizontally across the gel. Further strips were processed to obtain enzyme activity, protein distribution and pH gradient as indicated above.

In the digitonin-solubilised enzyme preparations, mannan synthase and galactosyltransferase activity peaks overlapped. The mannan synthase gave a broad peak at about pI 6.0, tailing towards the origin of the gel where a large proportion of the activity remained, apparently

unable to enter the gel. By contrast, all of the galactosyltransferase activity entered the gel, and gave a more symmetrical peak [pI 6.0] overlapping with that of the mannan synthase. There was a good correlation between the galactosyltransferase activity and a protein band with an apparent molecular weight [Mr] of about 50K. The corresponding [about 50K] band was identifiable on one-dimensional SDS gels of digitonin-solubilised enzyme. Such gels were therefore electroblotted blotted onto "Problott" [Applied Biosystems] membrane, and the excised blotted band was subjected to N-terminal sequencing. Repeated attempts gave no sequence, indicating that the protein was blocked to sequencing at the N-terminus. To obtain internal sequence data, the band was excised from one-dimensional SDS gels and subjected to digestion with endo-proteinase GluC followed by separation of product peptides on SDS gels [Cleveland *et al* (1977) J Biol Chem 252: 1102-1106]. The peptides were electroblotted and subjected to N-terminal sequencing to give internal sequence data from the 50K protein. When the sequence information obtained was compared with international database information, there was extremely high homology between the obtained sequences and those of membrane bound provicilin storage protein precursors. This indicated either that the about 50K band identified on 2-D gels was not the galactosyltransferase, or that the corresponding band excised from the one-dimensional SDS gels contained more than one protein, the vicilin-related protein predominating.

Extension of the above method for use with other detergents, and refinements giving more rapid galactosyltransferase localisation in IEF gels with higher precision

Following our observation that galactosyltransferase activity was retained in digitonin extracts, and could be assayed using mannohexaose as described above, other detergents which, unlike digitonin, had abolished mannan synthase activity almost entirely were investigated. All those tested [Triton X-100, NP-40, CHAPS, octyl glucoside] gave some retention of activity, but Triton X-100 and NP-40 gave very high retentions, approximately double that observed for digitonin. The properties, including transfer-specificity of the Triton-solubilised enzyme were compared with and found identical with those of the digitonin-solubilised enzyme. This allowed the IEF / SDS-PAGE separation described above to be carried out using Triton X-100 in place of digitonin. This gave greatly improved activity resolution and protein separation.

Also following our observation that manno-oligosaccharides would serve as acceptors for detergent-solubilised galactosyltransferase, polymeric galactomannans with low, medium and high galactose-substitution were tested as acceptors [locust bean, guar and fenugreek galactomannans respectively]. Locust bean galactomannan was an efficient acceptor, guar galactomannan was less efficient and fenugreek galactomannan was not an acceptor. When the labelled products of transfer of galactose residues to guar and locust bean gums were subjected to hydrolysis using the structure-sensitive endo-β-mannanase, the distribution of label in the fragment oligosaccharides was consistent with transfer to relatively unsubstituted regions of the mannan backbone.

It was found that the commercial agarose preparation ["IsoGel" - FMC Bio-products] sold for isoelectric focussing is an agarose-galactomannan blend. On enzymatic digesting a sample of the blend with the structure-sensitive endo-mannanase the "fingerprint" of galactomannan-derived oligosaccharides observed on TLC was consistent with a low-galactose galactomannan, probably locust bean gum. The presence of a low-galactose galactomannan in the IEF agarose gel offered the possibility of its use as an *in situ* acceptor for gel-separated galactosyltransferase, and the design of a new rapid, sensitive, highly resolving procedure for localising the enzyme activity. To localise galactosyltransferase activity in an IEF gel strip, the entire strip could be incubated in the presence of labelled UDP-Gal, whereby galactosyltransferase focussed within the strip would catalyse the transfer of labelled galactose residues to the galactomannan component of the separating gel. After thorough washing of the gel, any radioactivity remaining within it was a measure of and a localisation of galactosyltransferase activity.

In practice a complete gel strip [on GelBond] cut parallel to the direction of current flow was pre-incubated in strong buffer [200 mM Tris-HCl pH 7.5] for 10 min in order to bring it to the correct pH for galactosyltransferase assay. The whole strip was then incubated in a mixture containing 50 mM Tris-HCl pH 7.5, 10 mM MnCl2, 0.2% [w/v]Triton X-100 and 800 μM 14C-labelled UDP-Gal for 3 hours at 30°C. The strip was then fixed in 40% [v/v] methanol / 10% [v/v] acetic acid for 20 minutes and washed overnight in 40%methanol. This procedure removed virtually all unincorporated label, and retained the labelled galactomannan product within the gel. The following day the gel strip was cut into 2 mm strips perpendicular to the direction of current flow. Each strip was removed from the GelBond, transferred to a

microcentrifuge tube, washed once with hot [60°C, 10 min] and twice with cold [room temp, 20 min] 40% methanol, dissolved in 20 µl concentrated HCl, and subjected to liquid scintillation counting.

In the case of Triton X-100 gels analysed using the procedure described above for galactosyltransferase localisation, the galactosyltransferase focussed at about pI 6.5. Correlation of activity with protein bands on SDS gels as above gave excellent register of activity with protein at about Mr 50K. However in contrast to the digitonin gels, the protein at about 50K was resolved into two main components. The correlation between the galactosyltransferase activity and one of these two components [Mr 51K] was very close. Western immunoblotting showed that this protein did not cross react with an anti-vicilin poyclonal antiserum. The second major band at about 50K [Mr 49K] cross-reacted strongly with the antiserum, indicating that it was the provicilin storage protein precursor mentioned above [Fig. 2].

To purify a small quantity of the 51K protein and obtain protein sequence, the material focussing at pI 6.2 to 6.8 was excised from an entire Triton X-100 IEF gel and the gel sections were applied as the sample to an SDS-PAGE gel. After running, the gel was blotted onto Problott membrane, and the blot was stained lightly with Coomassie blue. The 51K and 49K bands were adequately separated, and the 51K band was excised carefully from the blot and subjected to N-terminal sequencing. Sequence was obtained. To obtain internal sequence information from the 51K protein IEF and SDS-PAGE was carried out as above. The gels were stained lightly, and the 51K band was excised and subjected to digestion in gel (Cleveland et al. 1977) with endoproteinase GluC. The product polypeptides were separated by SDS-PAGE, blotted and subjected to N-terminal sequencing. Only one of the polypeptides was present in sufficient quantity to give a sequence [Internal 1, Table 1]. Further sequence data was obtained by in-gel digestion with endoproteinase LysC, separation of the resultant peptides by HPLC and direct sequencing [Internal 2 and 3, Table 1]. All protein sequence data were compared with international protein databases, and there were no significant sequence homologies.

Table 1. N-terminal and internal sequence information from 51K possible galactomannan galactosyltransferase.

Identification	Sequencing data

[see text]

N-terminal	ATKFGSKN-S-PWL
Internal 1	GY-LEISKMYDKMGE-YD
Internal 2	FGFIHPNLLDK
Internal 3	SVSPLPFGYPAASP

Example 2. Acquisition of a cDNA sequence encoding the 51K probable galactosyltransferase protein.

Degenerate primers were designed to the amino acid sequence information acquired from the 51K probable galactosyltransferase protein:

These were:

GY-LEISKMYDKMGE-YD

5' AAGATGTATGACAAGATGGG 3' (sense primer GT3S4)
A C T A

5' CCCATCTTGTCATACATCTT 3' (antisense primer GT3A4)
T A G T

ATKFGSKN-S-PWL

5' GCIACIAAATTTGGIA 3' (sense primer NTP2S)
G C T

RNA was prepared from endosperms hand-isolated from developing fenugreek seeds during the early stages of galactomannan deposition [32-35 days after anthesis (Edwards et al. 1992)]. When 3'RACE PCR [Frohman M A, Martin G R (1989) *Rapid amplification of cDNA ends using nested primers*. Techniques 1: 165-170] was carried out using a degenerate primer [GT3S4] designed to an internal protein sequence, a c500 bp cDNA was amplified. When cloned and sequenced [Fig. 3] it was found to encode further amino acid sequence from the 51K protein, adjoining that used to design the degenerate primer, and all the other internal sequence information shown in Table 1.1. In a procedure incorporating elements of 5'RACE [Frohman and Martin, 1989], and PCR

amplification using degenerate primers, a c1000 bp cDNA was amplified. When cloned and partially sequenced from both ends it encoded at the 5' end all the N-terminal amino acid sequence in Table 1, it overlapped at the 3' end with the c500 bp clone to the extent of the primer, and it encoded further amino acid sequence from "internal 1" [Table 1]. The partial terminal sequences of the c1000 bp clone are shown in Fig. 4. To obtain a single cDNA encompassing the whole sequence, perfect primers were designed to the extreme 5' end of the c1000 bp clone and to the 3' untranslated region of the c500 bp clone. RT-PCR, carried out using a proof-reading thermostable DNA polymerase [Pfu - Stratagene], resulted in the amplification of a c1500 bp cDNA which was cloned and fully sequenced. The complete sequence, shown in Fig. 5, had an orf encoding a 438 amino acid polypeptide. The deduced molecular weight was 51281 Daltons, and the deduced pI was 6.646, in close agreement with the values observed for the Triton X-100 solubilised 51K protein.

In more detail:

Preparation of RNA from developing fenugreek endosperms. Seeds from pods harvested 32-35 days after anthesis were hand-dissected under aseptic conditions, and the endosperm tissue was dropped directly into liquid nitrogen. Endosperms from 100 seeds [weight approx 1 g] were then ground in a mortar and pestle with liquid nitrogen, and RNA was prepared essentially according to the procedure of Lopez-Gomez R and Gomez-Lim M A (1992) *A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango*. HortScience 27: 440-442. This method, which involves an extraction buffer containing 20% ethanol, circumvented problems associated with the dissolution of galactomannan in extraction buffers. RNA yields were typically about 50 µg.

Design of degenerate primers to amino acid sequence from the 51K protein. A degenerate primer was designed to the extreme N-terminal part of the 'N-terminal [Table 1.1]' amino acid sequence and designated NTP2S. A further degenerate primer pair [sense and antisense, designated GT3S4 and GT3A4 was designed to part of the 'internal 1 [Table 1.1]' sequence (see above).

Use of 3'RACE PCR to obtain a c500 bp clone. 3' Rapid amplification of cDNA ends [3' RACE] was carried out essentially according to Frohman and Martin (1989). First strand cDNA synthesis from fenugreek endosperm RNA was primed using the $(dT)_{17}$-R_i-R_o primer described by Frohman and Martin (1989), and PCR was carried out using the degenerate primer

GT3S4 and a T7 RNA polymerase promoter primer [5' TAATACGACTCACTATAGGG 3'] recognising part of the R_i-R_o sequence. The PCR reaction mixture comprised 5 µl first strand cDNA, 25 pmol T7 primer, 25 pmol GT3S4 primer, 0.01 µmol of each dNTP, 2.5 U Taq polymerase [Pharmacia] and 10x Taq polymerase buffer [5 µl; Pharmacia] in a total volume of 50 µl. To obtain a "hot start" the template, in 30 µl, was heated to 95°C for 7 min, and then held at 75°C whilst the remaining components were added. The complete mixture was heated at 50°C for 2 min, followed by 72°C for 5 min, subjected to 30 cycles of 94°C [1 min]- 50°C [1 min] - 72°C [1.5 min], and then held at 72°C for 15 min. Agarose gel electrophoresis of the PCR mixture gave a weak signal at c500bp. Reamplification by PCR using the same primers and conditions gave a very strong signal on gels at c500 bp. The remainder of the PCR reamplification mixture was purified [Hybaid Recovery DNA purification kit] and cloned into the commercial [Invitrogen] plasmid pCR 2.1 using the 3' A overhangs resulting from the action of Taq DNA polymerase. The cDNA fragment was subcloned and sequenced. The encoded amino acid sequence included further known amino acid sequence from the "internal 1" peptide used to design the degenerate primer, and sequences corresponding to the "internal 2" and "internal 3" sequences obtained directly from the 51K protein [Fig. 3].

Use of a modified 5'RACE PCR protocol to obtain a c1000bp clone. Initially 5' RACE was carried out essentially according to Frohman and Martin (1989). First strand cDNA synthesis from fenugreek RNA was primed using random hexamers, and polyA tailed at the 3' end using terminal transferase. Second strand synthesis was primed using the $(dT)_{17}$-R_i-R_o primer described by Frohman and Martin (1989) and PCR amplification was carried out using the T7 promoter primer described above and a perfect primer [5' CATTTCACCATAACGTTCACTCAC 3'designated GTPA3] designed to part of the sequence of the c500bp clone [Fig. 3]. The procedure of Frohman and Martin was modified by carrying out the second strand synthesis and PCR amplification in separate stages. In the first stage, "hot-started" as above, poly A tailed first strand cDNA [5 µl], $(dT)_{17}$-R_i-R_o primer [2.5 pmol], dNTP's [0.01 µmol each], with Taq polymerase [2.5 U] and Taq polymerase buffer [Pharmacia] were heated at 45°C for 2 min and then 72°C for 10 min. In the second stage, T7 primer and primer GTPA3 were added to the above mixture whilst it was held at 72°C. The combined mixture was then subjected to 30 cycles of 94°C [1 min] -50°C [1 min] - 72°C [1.5 min], and then held at 72°C for 15 min. This procedure resulted in the amplification of DNA covering a wide range of molecular sizes, which was purified free of primers and

low molecular weight products [Hybaid Recovery], and PCR amplified using the degenerate N-terminal primer NTP2S and the degenerate internal antisense primer GT3A4 [see above]. The PCR protocol [with "hot start" as above]comprised 30 cycles of 94°C [1 min] - 37°C [1 min] 72°C [2 min], with a final period at 72°C [15 min]. This resulted in the amplification of a c1000 bp cDNA which was excised from gels, purified [Hybaid Recovery] and cloned, subcloned and sequenced from both ends. The sequence encoded at the 5' end the full "N-terminal" [Table 1.1] sequence from the 51K protein and, at the 3' end, the part of the "internal 1" [Table 1.1] sequence used to design the primer plus all the other amino acids towards the N-terminus of the "internal 1" peptide [Fig. 4].

RT-PCR amplification of a single cDNA encoding the full protein sequence. Perfect primers were designed to the 5' terminus of the c1000 bp clone [5' GCGACGAAATTTGGTTCCAA 3', designated GTP5S] and to part of the 3' untranslated region of the c500 bp clone [5' GCTAATATCATCACCACCTTC 3', designated GTP6A], [Fig. 5] and RT-PCR was carried out on fenugreek endosperm RNA, using the proofreading Pfu [Stratagene] DNA polymerase. First strand synthesis was primed using the $(dT)_{17}$-R_i-R_o primer. The PCR mixture , ["hot-started" as above] comprised first strand cDNA template, GTP5S and GTP6A primers [25 pmol each], dNTP's [0.01 μmol each], Pfu DNA polymerase [2.5 U; Stratagene] and 10x Pfu buffer [5 μl; Stratagene] in a total volume of 50 μl. The mixture was held at 50°C for 2 min, then at 72°C for 10 min before being subjected to 30 cycles of 94°C [1 min] - 50°C [1 min] - 72°C [4 min] and held at 72°C for 15 min. This resulted in the amplification of a c1500 bp fragment which was excised from the gel and purified [Hybaid Recovery]. The 3' A overhangs necessary for ligation into the pCR2.1 vector were added subsequently in a reaction containing purified DNA, dATP [0.01 μmoles] Taq polymerase [2.5U] and Taq buffer [Pharmacia] in a volume of 50 μl, heated to 72°C for 10 min. The cDNA was then purified [Hybaid Recovery], sub-cloned and sequenced. The sequence, which contained an orf of 1314 bp, encompassed all known sequence from the c500bp and c1000 bp clones. It encoded a 438 amino acid protein, deduced molecular weight 51282 and deduced pI 6.646. The deduced protein sequence included all amino acid sequence data obtained from the 51K protein, and was clearly the cDNA sequence which encoded it.

Protein database searching gave no significant homology with the deduced sequence. Secondary structure predictions carried out using the neural network based algorithms of Rost B and Sander C (1993) J Mol

Biol 232: 584-599; Proteins (1994) 19: 55-72; Proteins (1994) 20: 216-226; and Rost B et al (1995) Prot Sci 4: 521-533 [Fig. 6A and 6B]. Using a method designed specifically for prediction of transmembrane helices [Rost B, Casadio R, Fariselli P, Sander C (1995) Prot Sci 4: 521-533] a single transmembrane helix near the N-terminus of the protein was predicted [Fig. 6B]. Such sequences serve to anchor proteins to membranes, and are typical of many Golgi membrane bound proteins, including several glycosyltransferases [Paulson J C and Colley K J (1989) *Glycosyltransferases. Structure, localization and control of cell type specific glycosylation* J Biol Chem 264: 17615-17618].

Example 3. Evidence that the 51K protein is the fenugreek galactomannan galactosyltransferase

To establish with certainty a functional link between the 51K protein and the galactomannan galactosyltransferase, a strategy was devised to insert the encoding DNA sequence into a micro-organism. Any expressed fenugreek galactosyltransferase activity would be easily identified. It was recognised that expression of the full-length DNA including the transmembrane helix "anchor" sequence might lead to the attachment of any expressed protein to cellular membranes of the host microorganism. Thus our strategy included the expression not only of the full length 51K protein but also of a truncated protein lacking the sequence from the N-terminus to just beyond the transmembrane helix. The truncated protein, if expressed, might be expected to be enzymatically active but not membrane-bound.

It was decided to attempt to insert the cDNA sequences in-frame into the genome of the methylotrophic yeast *Pichia pastoris* under the control of an alcohol oxidase [AOX] promoter and the yeast α secretion factor. *Pichia* constructs were obtained for both the full-length and the truncated sequence, and culture filtrates were assayed for the activity of the fenugreek galactosyltransferase using locust bean galactomannan [low galactose] as acceptor substrate. Controls [no insert] gave no activity, full-length constructs gave moderate levels of activity, and truncated constructs gave very high levels of activity [Table 2].

Table 2 Galactomannan galactosyltransferase activities in 10x concentrated 44 hour culture supernatants from *Pichia* transformants, in relation to the activity in a typical Triton X-100 extract of fenugreek

membranes [not concentrated]

Sample	Activity [$\mu mol.l^{-1}h^{-1}$]	Activity [Rel.to fenugreek membranes]
Triton X-100 extract [fenugreek membranes]	10.4	1
Supernatant, colony 8 [full length insert]	1.6	0.015
Supernatant, colony 23 [full length insert]	4.2	0.041
Supernatant, colony 27 [truncated insert]	94.9	0.91
Supernatant, colony 29 [truncated insert]	116	1.11
Supernatant pPIC9 transformation [no insert]	0.11	0.001

Fragmentation of the labelled galactomannan product of the reaction, separation of the labelled oligosaccharides by TLC and digital autoradiography gave a pattern of labelled galactomanno-oligosaccharides identical with those obtained using the detergent-solubilised galactomannan galactosyltransferase [Fig 7]. The type of galactosyltransferase activity present in the culture supernatants from the *Pichia* transformants was identical with that of the solubilised fenugreek galactosyltransferase, providing proof that the 51K protein encoded the fenugreek galactomannan galactosyltransferase. The levels of secreted activity were high. Full length constructs gave activities approaching those in typical detergent extracts [see above], whilst truncated constructs gave very much higher levels of activity. This indicated *either* that the presence of the membrane-anchoring helical domain of the full-length protein hampered expression and/or secretion, *or* that the modified protein lacking the membrane anchor had a higher specific activity under our in vitro assay conditions.

PCR amplification of cDNA encoding the complete protein sequence and a truncated sequence lacking the transmembrane helix, with sequence extensions permitting insertion of the sequences in-frame into the genome of Pichia pastoris under the control of an AOX promoter and the

yeast α-secretion factor.

Primers, designated GTEXP1S, GTEXP2S and GTEXP3A were designed and synthesised to allow amplification of the entire sequence and of a truncated sequence with sequence extensions allowing cloning in-frame in the multiple cloning site of the *Pichia* expression vector pPIC9 [Invitrogen], using the Xho1 and Not1 restriction sites:

Sense primer GTEXP1S:

5' GTA TCT CTC GAG AAA AGA GCG ACG AAA TTT GGT TCC AAA 3'

A T K F G S K -

Sense primer GTEXP2S:

5' GTA TCT CTC GAG AAA AGA AAC TCC AAC CCA AAA TTC AAC 3'

N S N P K F N -

(Xho1 sites are underlined)

- Y P A A S P .

3' ATG GGG CGA CGT AGT GGT ATT TCC CGC CGG CGC TTA ATT 5'

(Not1 site underlined)

Using plasmid DNA with the full-length c1500 bp sequence [Fig. 5] as template, primers GTEXP1S and GTEXP3A amplified a c1400 bp band which was purified from gels, digested with Not 1 and Xho 1, re-purified and cloned into pPIC9 which had been previously digested with the same restriction enzymes. Ampicillin-resistant clones were screened for the presence of inserts by PCR using gene-specific primers and a primer designed to part of the α-factor sequence on the vector . This primer confirmed not only that apparently correct inserts were present but also confirmed their orientations. Plasmid DNA prepared from positive clones was further checked for the presence of the correct inserts by digestion with Xho1 and Not 1. Primers GTEXP2S and GTEXP3A amplified a c 1300 bp fragment which was similarly treated.

Transformation of Pichia. The pPIC9 constructs with the full length and truncated sequences, pPIC9F and pPIC9T respectively, were each amplified, and samples of each plasmid DNA were linearised with Stu 1 [pPIC9FStu1, pPIC9TStu1]. Competent cells of *Pichia pastoris* GS115 were prepared and transformed using the EasyComp [Invitrogen] kit. Separate transformations were carried out using pPIC9FStu1, pPIC9TStu1 and Stu 1 linearised pPIC9 as control. In each case, putative positive transformants were selected on the basis of their ability to grow on

histidine-free medium as described in the Invitrogen *Pichia* expression kit manual. Putative positives were further screened by direct PCR amplification of colonies. Yeast cells were boiled for 10 min prior to the addition of the PCR ingredients.

Assay for galactomannan galactosyltransferase activity associated with Pichia transformants. Putative positive transformants, assumed to be Mut+ [fast growing] as would be expected from the restriction enzyme used to cleave the pPIC9 vectors before transformation [Invitrogen Pichia expression kit manual], were inoculated, using single colonies, into 10 ml of BMGY [no methanol] medium in 50 ml conical tubes and grown at 30°C with continuous rotatory shaking [200 rpm] for 24 hours [A600 about 2.7]. Cells were harvested by centrifugation. The supenatants were decanted and the cells resuspended in BMMY [containing methanol] medium to give an A600 value of 1.0. Samples [50 ml] were further cultured at 30°C for 70 hours, samples being withdrawn at 0, 20, 44 and 70hours. Methanol was added to 0.5% at every sampling.

All samples were centrifuged, and supernatants were collected, concentrated [x10] using Vivapore [Vivascience] membrane concentrators [7.5K cut-off], and assayed for galactosyltransferase activity using locust bean galactomannan as galactosylacceptor. The assays [100 µl] contained 50 µl concentrated supernatant, 25 mM Tris-HCl buffer pH 7.5, 2 mM MnCl2, 0.2% [w/v] locust bean galactomannan and 800 µM labelled UDP-Gal, and were incubated at 30°C for 2 hours. At the end of the incubation time glacial acetic acid [50 µl] was added and the mixture was heated at 100°C for 2 min. The galactomannan acceptor was precipitated by adding methanol to a final concentration of 70%, washed exhaustively with hot 70% methanol as described previously, and either subjected to liquid scintillation counting or fragmentation using the structure sensitive endo-mannanase from *A niger*. Concentrated supernatants from pPIC9 controls contained no activity, whilst those from full-length constructs contained low activity, and those from truncated constructs showed very high activity comparable with the activities present in detergent extracts from membrane preparations. Typical activity data are shown in Table 2. When labelled galactomannan products were digested with the *A niger* endo-β-mannanase, the only labelled products of the reaction were diagnostic galactomannan oligosaccharides [Fig. 7].

Example 4 - identification of a homologous sequence from developing guar (*Cyamopsis tetragonoloba* [L] Taub.) endosperms, and demonstration

that it encodes a galactomannan galactosyltransferase

On further database searching, the fenugreek galactosyltransferase showed limited homology at the protein level with several yeast sequences known or believed to be galactosyltransferases, notably MN10_YEAST (SWISS-PROT: P50108) and GM12_SCHPO (SWISSPROT: Q09174). Degenerate sense and antisense primers (GT5S1 and GT5A1 - Table 3) were designed, following the fenugreek galactosyltransferase sequence, to a short region of very high homology between all three sequences. This covered amino acids 190 - 210 of the fenugreek galactosyltransferase sequence.

Table 3. Primers used to obtain the guar sequence

GT5S1	5' GAG TGG ATI TGG TGG GTI GAC 3'	
	A T	
GT5A1	5' TCI ACC CAC CAI ATC CAT TC 3'	
	C	
GT5S4	5' AGG CAT GCA GAG AAA GTG AGT 3'	
GT5A4	5' ACT CAC TTT CTC TGC ATG CCT 3'	
GT5A5	5' TTT TCG TCC CAG TTT TTC AT 3'	
	C A A C	
GP1A	5' GGC GTT CGT TGG GAT CGT AT 3'	
GP2S	5' GTA TCA CAT TCA CTC ACT CC 3'	

RNA was prepared, as for fenugreek, from the developing endosperms of guar seeds during the early stages of galactomannan deposition (30 to 35 days after anthesis, Edwards et al. 1992). First strand cDNA was synthesised, as before, using the $(dT)_{17}$-R_iR_o primer (Frohman and Martin, 1989). When 3'RACE was carried out using this first strand cDNA, primer GT5S1, and the T7 RNA polymerase promoter, an 800 - 900 bp cDNA was amplified.

To test whether the 800 - 900 bp band was likely to be a homologue of fenugreek galactosyltransferase, PCR amplification was carried out using sense and antisense primers designed to the fenugreek galactosyltransferase sequence between amino acid 210 and the C-terminus of the protein, paired with GT5S1 and T7, using the purified 800- 900 bp cDNA as template. One pair (GT5S4 and GT5A4, Table 3) gave efficient amplification of cDNA bands. GT5S4 in combination with T7 and GT5S1 in combination with GT5A4 resulted in the amplification of bands of the sizes expected if the 800 - 900 bp band amplified by 3'RACE was a sequence homologous to the fenugreek galactosyltransferase.

A 5'RACE protocol (Frohman and Martin, 1989) was carried out, modified as described in Example 2. First strand cDNA reverse-transcribed from guar RNA was primed using random hexamers, and polyA tailed at the 3' end using terminal transferase. Second strand synthesis was primed using the $(dT)_{17}$-R_i-R_o primer. A first round of PCR amplification was carried out using the R_i primer (Frohman and Martin 1989) and GT5A4 (Table 3). Amplified cDNA was recovered (Hybaid Recovery) and used as template for a second round amplification, using the degenerate primer (NTP2S, as used in Example 2) designed to the N-terminal protein sequence of the fenugreek galactosyltransferase along with GT5A1. This resulted in the amplification of a 570 bp cDNA.

Alternative second round amplifications were attempted using primer R_i and antisense primers designed to the fenugreek galactosyltransferase sequence between amino acid 190 and the N-terminus. This was in order to amplify sequence extending 5' of the terminus of the fenugreek sequence. One of these primers (GT5A5, designed to amino acids 96 - 116, Table 3) resulted in the amplification of a 400 bp cDNA.

All of the above cDNA's were gel purified [Hybaid Recovery], cloned, and subcloned. Sequence data obtained from them was aligned to give a composite clone. Perfect sense and antisense primers were designed to sequences near the 5' end and the 3' end respectively of the composite sequence, and used in RT-PCR reactions using guar RNA as template and the Pfu proof-reading DNA polymerase (Stratagene). The combination GP2S and GP1A (Table 3) resulted in the amplification of a single c 1400 bp cDNA. This was gel purified, and ligated into the commercial plasmid pCR 2.1 TOPO (Invitrogen), and *E coli* cells (TOP 10F', Invitrogen)

were transformed with the ligation mixture, and transformants were amplified. Single colonies were PCR checked and a single positive colony was used to prepare plasmid DNA, which was used as template for the full sequencing of the c 1400 bp insert. A series of plasmid based and gene-specific sequencing primers was used. The base sequence of the insert is shown in Annex 2a.

The sequence showed a continuous open reading frame from the start to base 1326 (Annex 2b). Near the 5' end (base 24 onwards), the encoded protein sequence was closely similar to that at the the N-terminus of the fenugreek galactosyltransferase (ATKFGS in fenugreek, and AKFGS in guar). In guar, this sequence was immediately preceded by a methionine residue, which may represent the start of translation. On this assumption, the cDNA encoding the putative guar galactosyltransferase comprises 1305 bp and encodes a 435 amino acid protein. The fenugreek galactosyltransferase and the putative guar galactosyltransferase are aligned in Fig 1.

Clearly the two sequences are highly similar (77% similarity; differences highlighted in Fig. 1). As for the fenugreek galactosyltransferase (Example 2), secondary structure predictions [Rost B and Sander C (1993) J Mol Biol 232: 584-599; (1994) Proteins 19: 55-72; (1994) Proteins 20: 216-226; Rost B et al (1995) Prot Sci 4: 521-533] revealed a single, membrane-spanning helix near the N-terminus (underlined in Fig. 1). To establish whether or not the guar sequence was functionally a galactomannan galactosyltransferase, the full-length protein (residues 1 - 435, Fig. 1) and a truncated protein lacking the membrane spanning helix (residues 43 - 435, Fig 1) were separately overexpressed in *Pichia pastoris* as described for the fenugreek sequence in Example 2. The experimental strategy employed was exactly as in Example 2.

Pichia constucts were obtained for both the full-length and the truncated sequence, and culture filtrates were assayed for galactomannan galactosyltransferase activity exactly as described in Example 1. Culture supernatants from control transformants (no insert) and from transformants with full-length inserts did not contain measurable amounts of galactomannan galactosyltransferase activity, whereas supernatants from transformants with truncated inserts

contained significant levels of activity (7.99 ± 1.90 $\mu mol.l^{-1}.h^{-1}$; 6 independent clones; supernatants not concentrated). When labelled galactomannan products formed by the catalytic action of the expressed guar protein were digested with the structure-sensitive *A niger endo*-β-mannanase, the only digestion products were again oligosaccharides diagnostic of legume-seed galactomannan.

Example 5 - production of transgenic plants

Transgenic plants containing modified levels of the fenugreek or guar galactosyltransferase genes, or derivatives thereof, may be produced using methods known to those skilled in the art. Gene constructs will be expressed constitutively or in a tissue-specific manner in the seed or endosperm, potentially at a specific developmental stage. Constructs may include antisense versions of e.g. guar galactosyltransferase. Transgenic Guar plants may then be produced, for instance using methods analogous to those discussed in WO 97/20937. This will result in guar galactomannan with a higher man/gal ratio.

Example 5 - Foodstuffs comprising modified galactomannan

Modified galactomannans may be extracted from transgenic plants by methods analogous to those used in the art.

An ice cream based on the modified galactomannan may be provided as follows:

Ingredient	Amount
Galactomannan	0.35
Liquid sugar	15
Skimmed Milk (30% solids)	15.9
Butter fat	9
Espiron 300	5
MGp	0.3
Flavour	0.01
Colour	0.004
Water	to 100

A water ice may be provided as follows:

Ingredient	Amount

Galactomannan	0.1
Liquid sugar	15.7
Liquid dextrose	4
Citric acid	0.2
Flavour	2.6
Colour	0.0075
Water	to 100

Sequence Annexes

Annexe 1a: fenugreek cDNA sequence - Seq ID No 1

gcgacgaaatttggttccaaaaacaaatcctctccatggctctcaaatggttgcatcttcctcctaggtgc
aatgtcagc
tcttcttatgatttgggggctcaattccttcatcgctccaatcccaaactccaacccaaaattcaactcct
tcaccacca
aactcaaatccttaaacttcaccacaaacaccaactttgctggtcctgatttgttacatgacccttcagac
aaaaccttc
tatgatgatccagaaacatgttacaccatgatggacaaaccaatgaaaaattgggatgagaagcgtaaaga
atggctatt
tcatcatccctcattcgcggctggagcaaccgaaaagatacttgttataacgggttcacagccgacaaagt
gtgacaacc
ccatcggagaccacctttactaaggttctataaaaacaaggttgattattgtcgtatacacaaccacgac
ataatctac
aacaatgcattgttgcacccaaaaatggactcttactgggccaagtatcctatggttcgggccgcaatgtt
ggcccatcc
ggaagtagaatggatatggtgggtcgactctgatgccatctttaccgatatggaattcaagttaccgttat
ggcgttaca
aggatcacaaccttgtgattcatggttgggaagagttggttaagacagagcatagttggaccgggcttaac
gcgggtgtt
ttcttgatgaggaattgtcaatggtcgttggattttatggatgtttgggccagtatgggcccaaacagccc
ggaatacga
gaaatgggggagagacttagagaaacttttaagacaaaagtggtacgtgattcagatgatcagacggcgc
ttgcttact
tgatcgcgatgggagaggacaagtggacaaagaagatctatatggagaatgagtattattttgaagggtat
tggttagag
atttcaaagatgtatgataaaatgggtgagagatatgatgagatagaaaaaagagtggaagggttaaggag
gaggcatgc
agagaaagtgagtgaacgttatggtgaaatgagagaggagtatgttaagaatttaggggatatgagaagac
cttttatta
cacattttacagggtgccaaccttgtaatggtcatcataatccaatatatgctgcagatgattgctggaat
ggcatggag
agagctctcaattttgctgataatcaggtgttgcgcaagtttggtttcattcatccaaatctattggataa
gtctgtttc
tccattaccatttggataccccgctgcatcacca

Annex 1b: translation of the fenugreek cDNA sequence - Seq ID No 2

ATKFGSKNKSSPWLSNGCIFLLGAMSALLMIWGLNSFIAPIPNSNPKFNSFTTKLKSLNFTTNTNFAGPDL
LHDPSDKTFYDDPETCYTMMDKPMKNWDEKRKEWLFHHPSFAAGATEKILVITGSQPTKCDNPIGDHLLLR
FYKNKVDYCRIHNHDIIYNNALLHPKMDSYWAKYPMVRAAMLAHPEVEWIWWVDSDAIFTDMEFKLPLWRY
KDHNLVIHGWEELVKTEHSWTGLNAGVFLMRNCQWSLDFMDVWASMGPNSPEYEKWGERLRETFKTKVVRD
SDDQTALAYLIAMGEDKWTKKIYMENEYYFEGYWLEISKMYDKMGERYDEIEKRVEGLRRRHAEKVSERYG
EMREEYVKNLGDMRRPFITHFTGCQPCNGHHNPIYAADDCWNGMERALNFADNQVLRKFGFIHPNLLDKSV
SPLPFGYPAASP

Annexe 2a: guar cDNA sequence - Seq ID No 3

GTATCACATTCACTCACTCCCATGGCCAAATTTGGTTCCAGAAACAAATCCCCTAAATGGA
TCTCCAACGGTTGCTGCTTCCTCCTAGGAGCATTCACTGCTCTTCTTCTGCTCTGGGGTTTA
TGCTCCTTCATCATCCCCATCCCAAACACCGACCCCAAGCTCAACTCCGTCGCCACCAGTT
TGAGATCCCTTAACTTTCCCAAAAACCCCGCTGCCACCTTGCCTCCCAACTTGCAGCACGA
CCCTCCTGACACCACCTTCTACGACGACCCCGAAACCAGTTATACCATGGACAAACCAAT
GAAAAACTGGGACGAGAAGCGTAAGGAGTGGTTGCTGCATCATCCTTCGTTTGGCGCCGC
AGCACGCGATAAGATTCTCCTGGTGACAGGTTCTCAGCCGAAACGGTGCCATAACCCGAT
CGGCGACCACCTCCTGTTGCGGTTTTTCAAGAACAAGGTGGATTACTGCCGGCTGCACAAC
TACGACATAATTTACAACAACGCGCTTCTGCATCCTAAAATGAACTCTTATTGGGCCAAGT
ATCCAGTGATTCGGGCGGCGATGATGGCCCATCCGGAAGTGGAGTGGGTGTGGTGGGTGG
ACTCGGACGCGGTTTTCACGGACATGGAGTTCAAGCTTCCGTTAAAGCGTTATAAGAACC
ACAATCTGGTGGTTCACGGTTGGGAAGGATTGGTACGGTTGAACCATAGCTGGACGGGTC
TAAACGCGGGCGTATTCTTGATTCGGAATTGCCAGTGGTCGTTGGAGTTCATGGATGTG
TGGGTGAGCATGGGGCCACAGACTCCGGAATACGAGAAATGGGGGGAGAGGTTGAGAGAGA
CATTCAAGGACAAGGTGCTGCCTGATTCGGACGATCAGACGGCGCTGGCTTACCTGATCG
CGACGGATAATAAGGACACGTGGAGGGAGAAGATCTTCTTGGAGAGCGAGTACTACTTCG
AAGGGTACTGGCTGGAGATCGTGAAGACGTACGAGAACATAAGCGAGAGGTATGATGAG
GTGGAGAGGAAGGTGGAAGGGTTGAGGAGGAGGCATGCGGAAAAGGTGAGCGAGAAAT
ACGGTGCGATGAGGGAGGAGTATCTGAAGGACAACAAGAGGAGGCCCTTTATCACGCAC
TTTACTGGGTGTCAACCCTGTAATGGCCACCATAATCCTGCTTATAATGCTAATGATTGCT
GGAATGGCATGGAGAGGGCTCTTAATTTCGCTGATAATCAAATCTTGCGTACTTACGGTTA
TCACCGTCAAAATTTACTCGACAAGTCTGTTTCACCCTTACCTTTTGGTTACCCTGCTGCAT
AATAATGTACTACTACTGATAACGACAGTTATTTAAAATTTATTATACGATCCCAACGAAC
GCC

Annex 2b: translation of the guar cDNA sequence - Seq ID No 4

VSHSLTPMAKFGSRNKSPKWISNGCCFLLGAFTALLLLWGLCSFIIPIPNTDPKLNSVATSLRSLNFPKNP
AATLPPNLQHDPPDTTFYDDPETSYTMDKPMKNWDEKRKEWLLHHPSFGAAARDKILLVTGSQPKRCHNPI

GDHLLLRFFKNKVDYCRLHNYDIIYNNALLHPKMNSYWAKYPVIRAAMMAHPEVEWVWWVDSDAVFTDMEF
KLPLKRYKNHNLVVHGWEGLVRLNHSWTGLNAGVFLIRNCQWSLEFMDVWVSMGPQTPEYEKWGERLRETF
KDKVLPDSDDQTALAYLIATDNKDTWREKIFLESEYYFEGYWLEIVKTYENISERYDEVERKVEGLRRRHA
EKVSEKYGAMREEYLKDNKRRPFITHFTGCQPCNGHHNPAYNANDCWNGMERALNFADNQILRTYGYHRQN
LLDKSVSPLPFGYPAA..CTTTDNDSYLKFIIRSQRT

Claims

1 An isolated nucleic acid encoding a polypeptide which is capable of catalysing the biosynthesis of a complex non-cellulosic plant cell wall polysaccharide.

2 A nucleic acid as claimed in claim 1 wherein the polysaccharide is a hemicellulose.

3 A nucleic acid as claimed in claim 1 or claim 2 wherein the polypeptide is a glycosyltransferase.

4 A nucleic acid as claimed in claim 3 wherein the polypeptide is a galactosyltransferase.

5 A nucleic acid as claimed in any one of claims 2 to 4 wherein the polysaccharide is galactomannan.

6 A nucleic acid as claimed in any one of the preceding claims having a sequence comprising Seq ID No 1 or is degeneratively equivalent thereto.

7 A nucleic acid as claimed in any one of the preceding claims having a sequence comprising Seq ID No 3 or is degeneratively equivalent thereto.

8 A nucleic acid as claimed in any one of claims 1 to 5 which is a homologous variant of Seq ID No 1.

9 A nucleic acid as claimed in any one of claims 1 to 5 which is a homologous variant of Seq ID No 3.

10 A nucleic acid as claimed in claim 8 or claim 9 wherein the variant is an allelic or pseudoallelic variant of Seq ID No 1 or Seq ID No 3.

11 A nucleic acid as claimed in claim 8 having a sequence which is a derivative of Seq ID No 1 by way of addition, insertion, deletion or substitution of one or more nucleotides and which encodes a polypeptide having altered activity with respect to Seq ID No 2.

12 A nucleic acid as claimed in claim 8 wherein the derivative encodes a functional portion of Seq ID No 2.

13 A nucleic acid comprising at least 15 nucleotides having a sequence comprising, or being degeneratively equivalent to, part of Seq ID No 1.

14 A nucleic acid as claimed in claim 9 having a sequence which is a derivative of Seq ID No 3 by way of addition, insertion, deletion or substitution of one or more nucleotides and which encodes a polypeptide having altered activity with respect to Seq ID No 4.

15 A nucleic acid as claimed in claim 14 wherein the derivative encodes a functional portion of Seq ID No 2.

16 A nucleic acid comprising at least 15 nucleotides having a sequence comprising, or being degeneratively equivalent to, part of Seq ID No 3.

17 A nucleic acid which is complementary to the nucleic acid of any one of claims 6 to 16.

18 A method for identifying or cloning a glycosyltransferase from a plant which method employs a nucleic acid molecule having a nucleotide sequence comprising, or complementary to, all or part of Seq ID No 1 or Seq ID No 3, or a derivative of either.

19 A method as claimed in claim 18 comprising the step of searching a data-base to find sequences which are homologous to Seq ID No 1 or Seq ID No 3.

20 A method as claimed in claim 18 comprising the steps of:
(a) providing a preparation of nucleic acid,
(b) providing a nucleic acid molecule having a nucleotide sequence comprising, or complementary to, all or part of the nucleic acid of claim 6 or claim 7,
(c) contacting nucleic acid in said preparation with said nucleic acid molecule under conditions for hybridisation of said nucleic acid molecule to any said gene or homologue in said preparation, and
(d) identifying said gene or homologue if present by its hybridisation with said nucleic acid molecule.

21 A method as claimed in claim 20 wherein the hybridisation conditions are selected to allow the identification of sequences having about 70% or more sequence identity with the nucleic acid molecule.

22 A method as claimed in claim 18 comprising use of two primers to amplify a nucleic acid encoding a glycosyltransferase,at least one of the primers having a sequence comprising, or complementary to part of Seq ID No 1 or Seq ID No 3 or a derivative of either.

23 A method as claimed in claim 22 comprising the steps of:
(a) providing a preparation of plant nucleic acid,
(b) providing a pair of nucleic acid molecule primers suitable for PCR, at least one of the primers having a sequence comprising, or complementary to part of the nucleic acid of claim 6 or claim 7,
(c) contacting nucleic acid in said preparation with said primers under conditions for performance of PCR,
(d) performing PCR and determining the presence or absence of an amplified PCR product.

24 A nucleic acid molecule for use as a probe or primer in the method of any one of claims 20 to 23, said molecule having a sequence comprising, or being complementary to, part of the nucleic acid of claim 6 or claim 7.

25 A recombinant vector comprising either the nucleic acid of any one of claims 1 to 17.

26 A vector as claimed in claim 25 which is capable of replicating in a suitable host.

27 A vector as claimed in claim 25 or claim 26 wherein the nucleic acid is operably linked to a promoter or other regulatory element for transcription in a host cell

28 A vector as claimed in claim 27 further comprising any one or more of the following: a terminator sequence; a polyadenylation sequence; an enhancer sequence; a marker gene.

29 A vector as claimed in claim 27 or claim 28 wherein the promoter is an inducible promoter.

30 A vector as claimed in any one of claims 25 to 29 which is a plant vector.

31 A vector as claimed in claim 30 comprising a selectable genetic marker which confers a selectable phenotype selected from: resistance to antibiotics or herbicides.

32 A method comprising the step of introducing a vector as claimed in any one of claims 26 to 31 into a cell.

33 A method for transforming a plant cell, comprising a method as claimed in claim 32, and further comprising the step of causing or allowing recombination between the vector and the plant cell genome to introduce the nucleic acid into the genome.

34 A host cell comprising a vector as claimed in any one of claims 26 to 31.

35 A host cell transformed with a vector as claimed in any one of claims 26 to 31.

36 A host cell as claimed in claim 34 or claim 35 which is a plant cell.

37 A host cell as claimed in claim 36 which is in a plant.

38 A method for producing a transgenic plant comprising a method as claimed in claim 33 and further comprising the step of regenerating a plant from the transformed cell.

39 A plant comprising the cell of claim 36 or claim 37.

40 A plant as claimed in claim 39 produced by the method of claim 38.

41 A plant as claimed in claim 40 which is an endospermic legume.

42 A plant which is the progeny of a plant as claimed in claim 40 or claim 41, and comprising the cell of claim 36 or claim 37.

43 A part or propagule of the plant of any one of claims 39 to 42.

44 A polypeptide encoded by the nucleic acid of any one of claims 1 to 16.

45 A method of producing a polypeptide comprising the step of causing or allowing the expression from a nucleic acid of any one of claims 1 to 16 in a suitable host cell.

46 A composition comprising the polypeptide of claim 44.

47 An antibody or fragment thereof, or a polypeptide comprising the antigen-binding domain of the antibody, capable of specifically binding the polypeptide of claim 44.

48 A method of producing the antibody or fragment as claimed in claim 47 comprising the step of immunising a mammal with a polypeptide according to claim 44.

49 A method of identifying and/or isolating a glycosyltransferase comprising the step of screening candidate polypeptides with a polypeptide comprising the antigen-binding domain of the antibody of claim 47.

50 A method for the *in vitro* synthesis of a polysaccharide comprising the use of the polypeptide of claim 44.

51 A method for altering the quality or quantity of a polysaccharide in a host cell by influencing the glycosyltransferase activity in that cell, the method comprising use of any one or more of the following: all or part of the nucleic acid of any one of claims 1 to 16; the polypeptide of claim 44; the antibody or fragment or polypeptide comprising the antigen-binding site thereof of claim 47.

52 A method as claimed in claim 51 wherein the polysaccharide is a complex non-cellulosic plant cell wall polysaccharide.

53 A method as claimed in claim 51 or claim 52 wherein the quality altered is galactose composition of the polysaccharide.

54 A method as claimed in claim 53 wherein the quality altered is the mannose:galactose ratio in a mannose/galactose containing polysaccharide in the cell.

55 A method as claimed in any one of claims 51 to 54 comprising the step of causing or allowing expression of a nucleic acid according to any one of claims 1 to 17 within the cell.

56 A method as claimed in any one of claims 51 to 55 comprising reducing the glycosyltransferase activity in the cell.

57 A method as claimed in claim 56 comprising the step of causing or allowing the transcription of part of the nucleic acid of any one of claims 1 to 16 in the cell such as to co-suppress the expression of an endogenous glycosyltransferase.

58 A method as claimed in claim 56 comprising the step of causing or allowing the transcription of nucleic acid of claim 17 in the cell.

59 A method as claimed in claim 56 comprising the step of causing or allowing the expression of a polypeptide comprising the antigen-binding domain of the antibody of claim 47.

60 A method as claimed in any one of claims 51 to 59 wherein the cell is a plant cell.

61 A method as claimed in claim 60 wherein the plant cell is part of a plant.

62 A method as claimed in any one of claims 51 to 61 wherein the glycosyltransferase is a galactosyltransferase.

63 A complex non-cellulosic plant cell wall polysaccharide the quality of which has been altered in accordance with the method of claim 61 or claim 62.

64 A plant product derived from any one of the plants of claims 39 to 42 or the plant cell of claims 36 or claim 37, said product comprising a complex non-cellulosic plant cell wall polysaccharide of claim 63.

65 A commodity comprising the altered cell wall storage polysaccharide of claim 63.

66 A commodity as claimed in claim 65 which is selected from: a human or animal foodstuff; a cosmetic.

67 A foodstuff as claimed in claim 66 which is a frozen food product.

68 A frozen food product as claimed in claim 67 which is an ice cream or water ice.

1/11

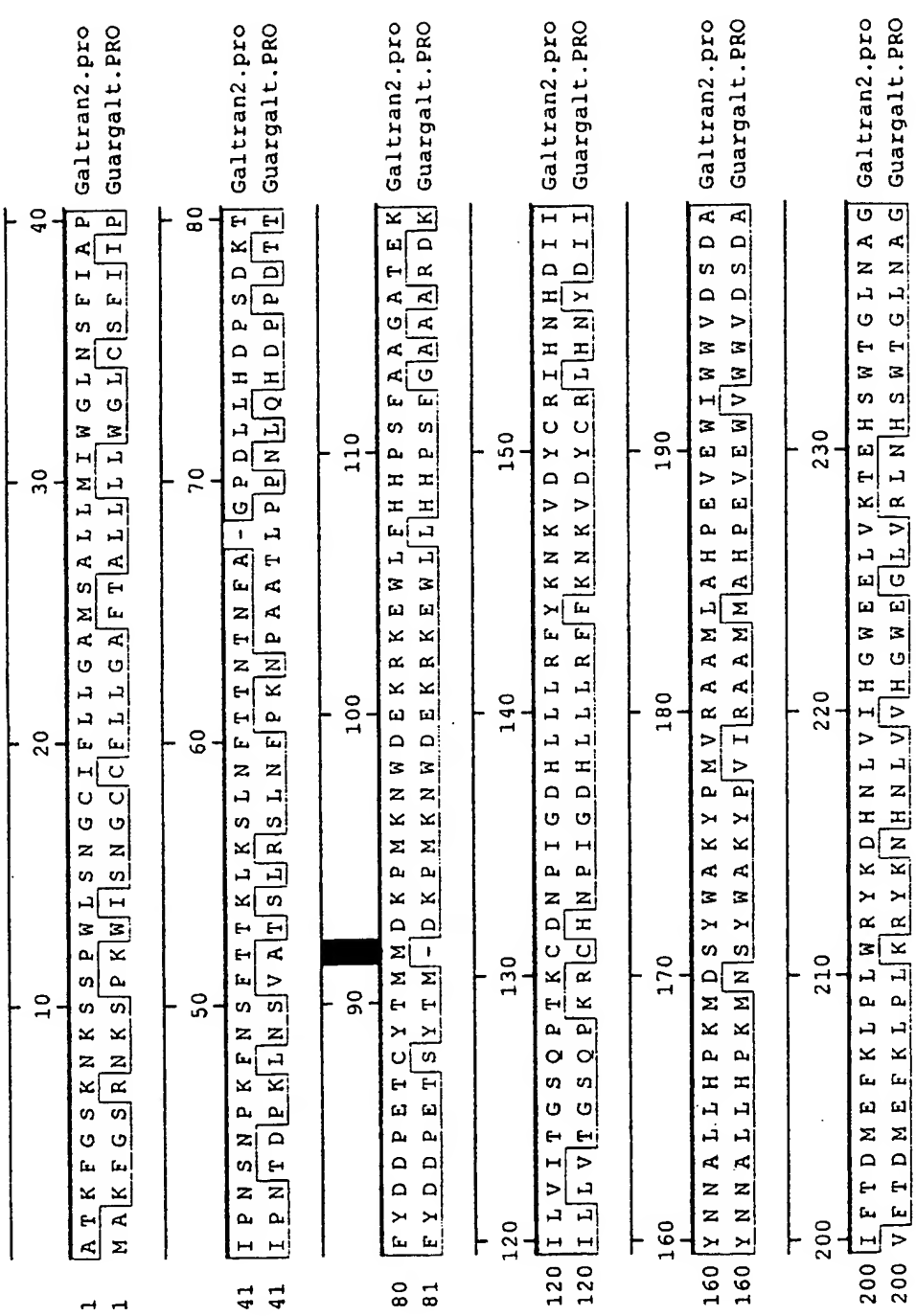

Fig. 1 (part 1 of 2)

2/11

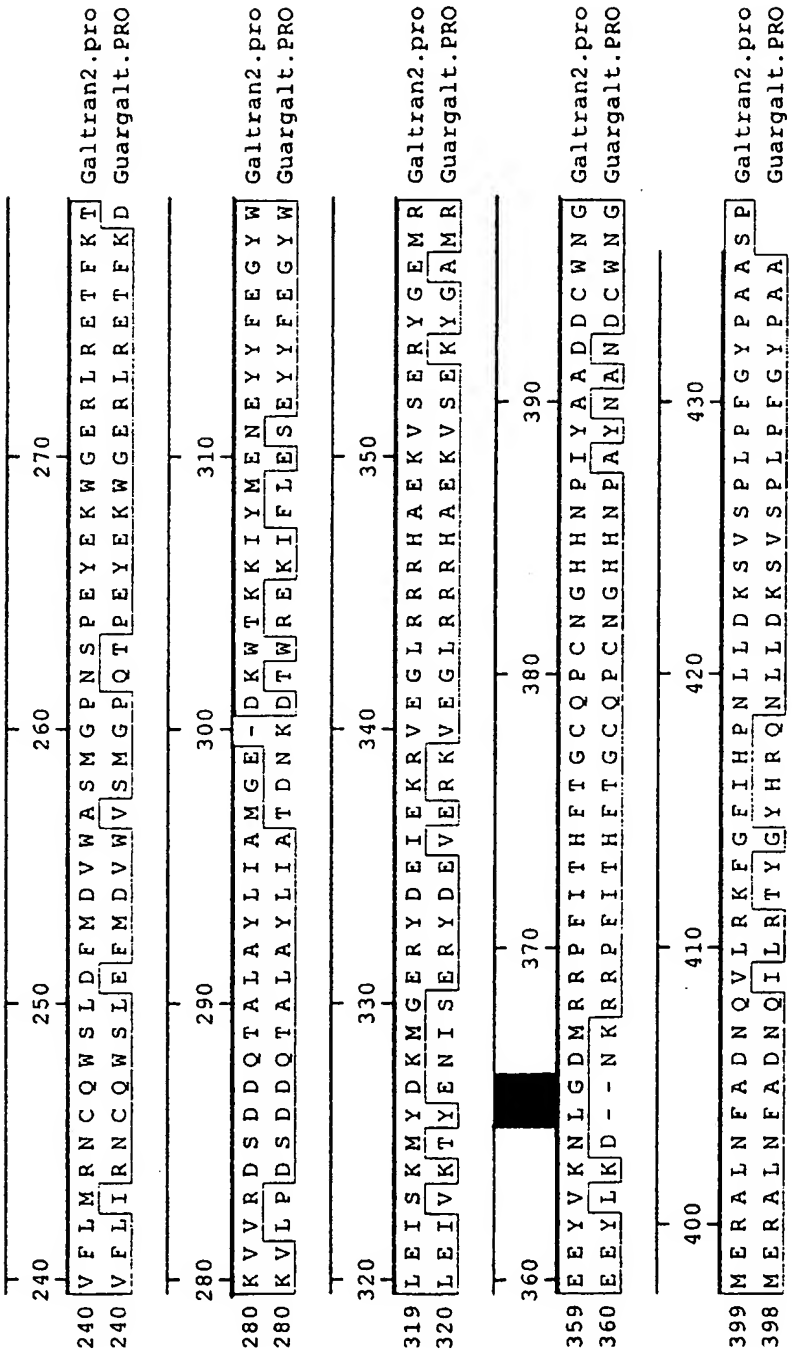

Fig. 1 (part 2 of 2)

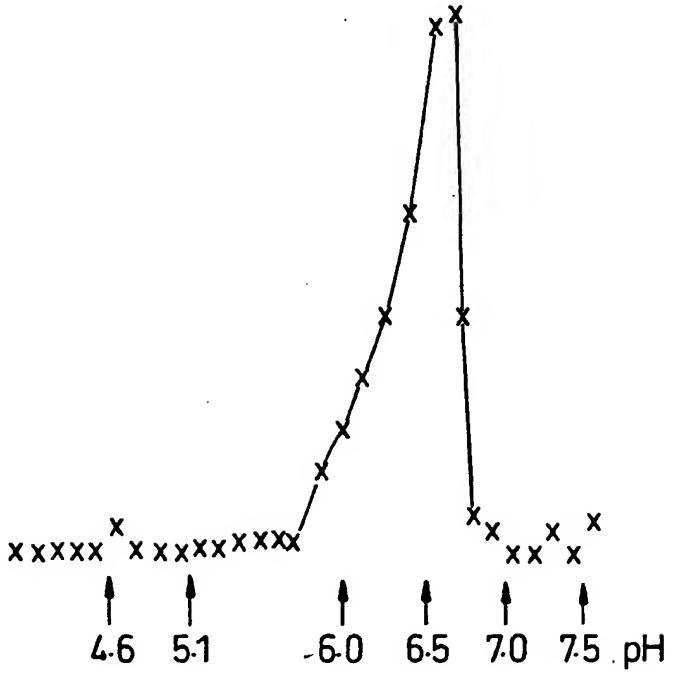

Fig. 2A

Fig. 2B

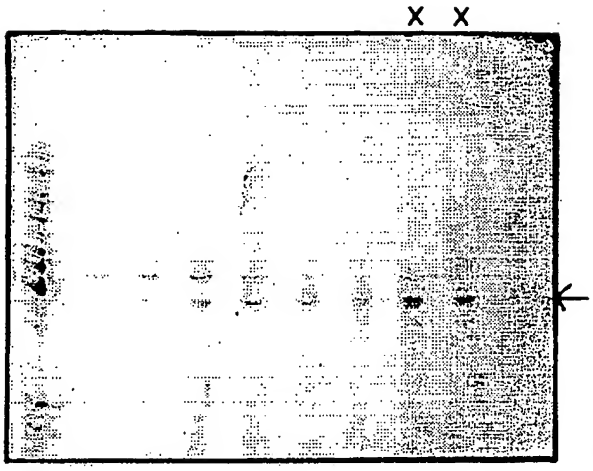

Fig. 2C

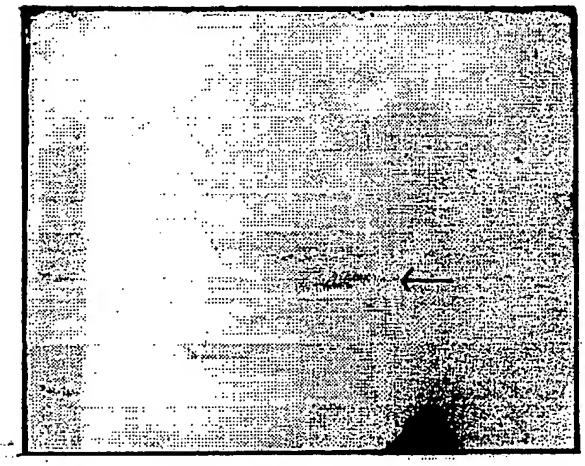

Fig. 2D

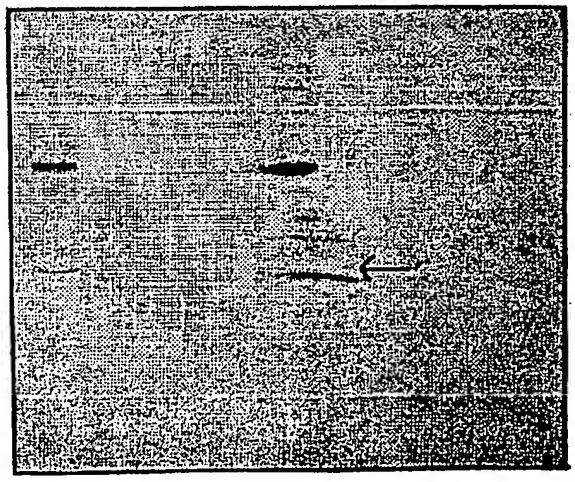

Fig. 2E

```
AAAATGTACGACAAAATGGGTGAGAGATATGATGAGATAGAAAAAAGAGTGGAAGGGTTAAGGAGGAGGCATGCASAGAAAGTGAGTGAA
TTTTACATGCTGTTTTACCCACTCTCTATACTACTCTTTTTTCTCACCTTCCCAATTCCTCCTCCGTACGTSTCTTTCACTCACTT

  K  M  Y  D  K  M  G  G  E  R  Y  D  E  I  E  K  K  R  V  E  G  L  R  R  R  H  A  ?  K  V  S  E

CGTTATGGTGAAATGAGAGAGGAGTATGTTAAGAATTTAGGGGATATGAGAAGACCTTTTATTACACATTTTACAGGGTGCCAACCTTGT
GCAATACCACTTTACTCTCTCCTCATACAATTCTTAAATCCCCTATACTCTTCTGGAAAATAATGTGTAAAATGTCCCACGGTTGGAACA

  R  Y  G  E  M  R  E  E  Y  Y  V  K  N  L  G  D  M  R  R  P  F  I  T  H  F  T  G  C  Q  P  C

AATGGTCATCATAATCCAATATATGCTGCAGATGATTGCTGGAACGGCATGGAGAGAGCTCTCAATTTTGCTGATAATCAGGTGTTGCGC
TTACCAGTAGTATTAGGTTATATACGACGTCTACTAACGACCTTGCCGTACCTCTCTCGAGAGTTAAAACGACTATTAGTCCACAACGCG

  N  G  H  H  N  P  I  Y  A  A  D  D  C  W  N  G  M  E  R  A  L  N  F  A  D  N  Q  V  L  R

AAGTTTGGTTTCATTCATCCAAATCTATTGGATAAGTCTGTTTYTCCATTACCATTTGGATACCCCGCTGCATCACCATAAATATATTAT
TTCAAACCAAAGTAAGTAGGTTTAGATAACCTATTCAGACAAARAGGTAATGGTAAACCTATGGGGCGACGTAGTGGTATTTATATAATA

  K  F  G  F  I  H  P  N  L  L  D  K  S  V  ?  P  L  P  F  G  Y  P  A  A  S  P  .  I  Y  Y

AACCTGCAGGGGTAAATTATATAATAGTAATTGTTATGATGATTYTTCTGTCAATAATAATCAAAATAATGAAGGTGGTGATGATATTAGCA
TTGGACGTCCCCATTTAATATTATCATTAACAATACTACTAARAAGACAGTTATTATTAGTTTTATTACTTCCACCACTACTATAATCGT

  N  L  Q  G  .  I  I  I  V  I  V  M  M  I  ?  L  S  I  I  I  K  I  M  K  V  V  M  I  L  A

ATTAATAAAAAAAAAAAAAAAAA
TAATTATTTTTTTTTTTTTTTTT

  I  N  K  K  K  K  K  K
```

Fig. 3

GCGACGAAATTTGGTTCCAAAAACAAATCCTCTCCATGGCTCTCAAATGGTTGCATCTTCCTCCTAGGTGCAATGTCAGCTCTTCTTATG
CGCTGCTTTAAACCAAGGTTTTTGTTTAGGAGAGGTACCGAGAGTTTACCAACGTAGAAGGAGGATCCACGTTACAGTCGAGAAGAATAC

A T K F G S K N K S S P W L S N G C I F L L G A M S A L L M

ATTTGGGGGCTCAATTCCTTCATCGCTCCAATCCCAAACTCCAACCCAAAATTCAACTCCATCACCGCCAAACTCAAATCCTTAAACTTC
TAAACCCCCGAGTTAAGGAAGTAGCGAGGTTTGAGGTTGGGTTTTAAGTTGAGGTAGTGGCGGTTTGAGTTTAGGAATTTGAAG

I W G L N S F I A P I P N S N P K F N S I T A K L K S L N F

ACCACAAACACCAACTTTGCTGGTCCTGATTTGTTACATGACCCTTCAGACAAAACCTTCTATGATGATCCAGAACATGTTACACCATG
TGGTGTTTGTGGTTGAAACGACCAGGACTAAACAATGTACTGGGAAGTCTGTTTTTGGAAGATACTACTAGGTCTTTGTACAATGTGGTAC

T T N T N F A G P D L L H D P S D K T F Y D D P E T C Y T M

ATGGACAAACCAATGAAAAATTGGGATGAGAAGCGTAAAGAATGGCTATTTCATCATCCCTCATTCGCG
TACCTGTTTGGTTACTTTTTAACCCTACTCTTCGCATTTCTTACCGATAAAGTAGTAGGGAGTAAGCGC

M D K P M K N W D E K R K E W L F H H P S F A

Fig. 4 (part 1 of 2)

TTTACCGATATGGAATTCAAGTTACCGTTATGGCGTTACAAGGATCACAACCTTGTGATTCATGGTTGGGAAGAGTTGGTTAAGAAAGAG
AAATGGCTATACCTTAAGTTCAATGGCAATACCGCAATGTTCCTAGTGTTGGAACACTAAGTACCAACCCTTCTCAACCAATTCTTTCTC

F T D M E F K L P L W R Y K D H N L V I H G W E E L V K K E

CATAGTTGGACCGGGCTTAACGCGGGTGTTTCTTGATGAGGAATTGTCAATGGTCGTTGGATTTTATGGATGTTTGGGCCAGTATGGGC
GTATCAACCTGGCCCGAATTGCGCCCACAAAGAACTACTCCTTAACAGTTACCAGCAACCTAAAATACCTACAAACCCGGTCATACCCG

H S W T G L N A G V F L M R N C Q W S L D F M D V W A S M G

CCAAACAGCCCCGAATACGAGAAATGGGGGGAGAGACTTTAAGACAAAGTGGTACGTGATTCAGATGATCAGACGGCG
GGTTTGTCGGGCCTTATGCTCTTTACCCCCCTCTCTGAATCTCTTTGAAAATTCTGTTTTCACCATGCACTAAGTCTACTAGTCTGCCGC

P N S P E Y E K W G E R L R E T F K T K V V R D S D D Q T A

CTTGCTTACTTCGATCGCGATGGGAGAGGACAAGTGGACAAAGAAGATCTATATGGAGAATGAGTATTATTTTGAAGGGTATTGGTTAGAG
GAACGAATGAACTAGCGCTACCCTCTCCTGTTCACCTGTTTCTTCTAGATATACCTCTTACTCATAATAAACTTCCCATAACCAATCTC

L A Y L I A M G E D K W T K K I Y M E N E Y Y F E *G Y W L E*

ATTTCAAAGATGTACGACAAGATGGG
TAAAGTTTCTACATGCTGTTCTACCC

I S K M Y D K M G

Fig. 4 (part 2 of 2)

```
GCGACGAAATTTGGTTCCAAAAACAAATCCTCTCCATGGCTCTCAAATGGTTGCATCTTCCTCCTAGGTGCAATGTCAGCTCTTCTTATG 90
CGCTGCTTTAAACCAAGGTTTTTGTTTAGGAGAGGTACCGAGAGTTTACCAACGTAGAAGGAGGATCCACGTTACAGTCGAGAAGAATAC
 A  T  K  F  G  S  K  N  K  S  S  P  W  L  S  N  G  C  I  F  L  L  G  A  M  S  A  L  L  M

ATTTGGGGGCTCAATTCCTTCATCGCTCCAATCCCAAACTCCAACCCAAAATTCAACTCCTTCACCACCAAACTCAAATCCTTAAACTTC 180
TAAACCCCCGAGTTAAGGAAGTAGCGAGGTTAGGGTTTGAGGTTGGGTTTTAAGTTGAGGAAGTGGTGGTTTGAGTTTAGGAATTTGAAG
 I  W  G  L  N  S  F  I  A  P  I  P  N  S  N  P  K  F  N  S  F  T  T  K  L  K  S  L  N  F

ACCACAAACACCAACTTTGCTGGTCCTGATTTGTTACATGACCCTTCAGACAAAACCTTCTATGATGATCCAGAAACATGTTACACCATG 270
TGGTGTTTGTGGTTGAAACGACCAGGACTAAACAATGTACTGGGAAGTCTGTTTTGGAAGATACTACTAGGTCTTTGTACAATGTGGTAC
 T  T  N  T  N  F  A  G  P  D  L  L  H  D  P  S  D  K  T  F  Y  D  D  P  E  T  C  Y  T  M

ATGGACAAACCAATGAAAAATTGGGATGAGAAGCGTAAAGAATGGCTATTTCATCATCCCTCATTCGCGGCTGGAGCAACCGAAAAGATA 360
TACCTGTTTGGTTACTTTTTAACCCTACTCTTCGCATTTCTTACCGATAAAGTAGTAGGGAGTAAGCGCCGACCTCGTTGGCTTTTCTAT
 M  D  K  P  M  K  N  W  D  E  K  R  K  E  W  L  F  H  H  P  S  F  A  A  G  A  T  E  K  I

CTTGTTATAACGGGTTCACAGCCGACAAAGTGTGACAACCCCATCGGAGACCACCTTTTACTAAGGTTCTATAAAAACAAGGTTGATTAT 450
GAACAATATTGCCCAAGTGTCGGCTGTTTCACACTGTTGGGGTAGCCTCTGGTGGAAAATGATTCCAAGATATTTTTGTTCCAACTAATA
 L  V  I  T  G  S  Q  P  T  K  C  D  N  P  I  G  D  H  L  L  L  R  F  Y  K  N  K  V  D  Y

TGTCGTATACACAACCACGACATAATCTACAACAATGCATTGTTGCACCCAAAAATGGACTCTTACTGGGCCAAGTATCCTATGGTTCGG 540
ACAGCATATGTGTTGGTGCTGTATTAGATGTTGTTACGTAACAACGTGGGTTTTTACCTGAGAATGACCCGGTTCATAGGATACCAAGCC
 C  R  I  H  N  H  D  I  I  Y  N  N  A  L  L  H  P  K  M  D  S  Y  W  A  K  Y  P  M  V  R

GCCGCAATGTTGGCCCATCCGGAAGTAGAATGGATATGGTGGGTCGACTCTGATGCCATCTTTACCGATATGGAATTCAAGTTACCGTTA 630
CGGCGTTACAACCGGGTAGGCCTTCATCTTACCTATACCACCCAGCTGAGACTACGGTAGAAATGGCTATACCTTAAGTTCAATGGCAAT
 A  A  M  L  A  H  P  E  V  E  W  I  W  W  V  D  S  D  A  I  F  T  D  M  E  F  K  L  P  L

TGGCGTTACAAGGATCACAACCTTGTGATTCATGGTTGGGAAGAGTTGGTTAAGACAGAGCATAGTTGGACCGGGCTTAACGCGGGTGTT 720
ACCGCAATGTTCCTAGTGTTGGAACACTAAGTACCAACCCTTCTCAACCAATTCTGTCTCGTATCAACCTGGCCCGAATTGCGCCCACAA
 W  R  Y  K  D  H  N  L  V  I  H  G  W  E  E  L  V  K  T  E  H  S  W  T  G  L  N  A  G  V

TTCTTGATGAGGAATTGTCAATGGTCGTTGGATTTTATGGATGTTTGGGCCAGTATGGGCCCAAACAGCCCGGAATACGAGAAATGGGGG 810
AAGAACTACTCCTTAACAGTTACCAGCAACCTAAAATACCTACAAACCCGGTCATACCCGGGTTTGTCGGGCCTTATGCTCTTTACCCCC
 F  L  M  R  N  C  Q  W  S  L  D  F  M  D  V  W  A  S  M  G  P  N  S  P  E  Y  E  K  W  G

GAGAGACTTAGAGAAACTTTTAAGACAAAAGTGGTACGTGATTCAGATGATCAGACGGCGCTTGCTTACTTGATCGCGATGGGAGAGGAC 900
CTCTCTGAATCTCTTTGAAAATTCTGTTTTCACCATGCACTAAGTCTACTAGTCTGCCGCGAACGAATGAACTAGCGCTACCCTCTCCTG
 E  R  L  R  E  T  F  K  T  K  V  V  R  D  S  D  D  Q  T  A  L  A  Y  L  I  A  M  G  E  D

AAGTGGACAAAGAAGATCTATATGGAGAATGAGTATTATTTTGAAGGGTATTGGTTAGAGATTTCAAAGATGTATGATAAAATGGGTGAG 990
TTCACCTGTTTCTTCTAGATATACCTCTTACTCATAATAAAACTTCCCATAACCAATCTCTAAAGTTTCTACATACTATTTTACCCACTC
 K  W  T  K  K  I  Y  M  E  N  E  Y  Y  F  E  G  Y  W  L  E  I  S  K  M  Y  D  K  M  G  E

AGATATGATGAGATAGAAAAAAGAGTGGAAGGGTTAAGGAGGAGGCATGCAGAGAAAGTGAGTGAACGTTATGGTGAAATGAGAGAGGAG 1080
TCTATACTACTCTATCTTTTTTCTCACCTTCCCAATTCCTCCTCCGTACGTCTCTTTCACTCACTTGCAATACCACTTTACTCTCTCCTC
 R  Y  D  E  I  E  K  R  V  E  G  L  R  R  R  H  A  E  K  V  S  E  R  Y  G  E  M  R  E  E

TATGTTAAGAATTTAGGGGATATGAGAAGACCTTTTATTACACATTTTACAGGGTGCCAACCTTGTAATGGTCATCATAATCCAATATAT 1170
ATACAATTCTTAAATCCCCTATACTCTTCTGGAAAATAATGTGTAAAATGTCCCACGGTTGGAACATTACCAGTAGTATTAGGTTATATA
 Y  V  K  N  L  G  D  M  R  R  P  F  I  T  H  F  T  G  C  Q  P  C  N  G  H  H  N  P  I  Y

GCTGCAGATGATTGCTGGAATGGCATGGAGAGAGCTCTCAATTTTGCTGATAATCAGGTGTTGCGCAAGTTTGGTTTCATTCATCCAAAT 1260
CGACGTCTACTAACGACCTTACCGTACCTCTCTCGAGAGTTAAAACGACTATTAGTCCACAACGCGTTCAAACCAAAGTAAGTAGGTTTA
 A  A  D  D  C  W  N  G  M  E  R  A  L  N  F  A  D  N  Q  V  L  R  K  F  G  F  I  H  P  N

CTATTGGATAAGTCTGTTTCTCCATTACCATTTGGATACCCCGCTGCATCACCATAAATATATTATAACCTGCAGGGGTAAATTATAATA 1350
GATAACCTATTCAGACAAAGAGGTAATGGTAAACCTATGGGGCGACGTAGTGGTATTTATATAATATTGGACGTCCCCATTTAATATTAT
 L  L  D  K  S  V  S  P  L  P  F  G  Y  P  A  A  S  P  .  I  Y  Y  N  L  Q  G  .  I  I  I

GTAATTGTTATGATGATTCTTCTGTCAATAATAATCAAAATAATGAAGGTGGTGATGATATTAGC 1415
CATTAACAATACTACTAAGAAGACAGTTATTATTAGTTTTATTACTTCCACCACTACTATAATCG
 V  I  V  M  M  I  L  L  S  I  I  I  K  I  M  K  V  V  M  I  L  A
```

Fig. 5

9/11

```
         ....,....1....,....2....,....3....,....4....,....5....,....6
AA       |ATKFGSKNKSSPWLSNGCIFLLGAMSALLMIWGLNSFIAPIPNSNPKFNSFTTKLKSLNF|
PHD sec  |               HHHHHHHHHHHHHHHHHHHHHHH
  HHH    |
Rel sec  |99788899889641235899999999999999887851699999978885424221012 2|

         ....,....7....,....8....,....9....,....10...,....11...,....12
AA       |TTNTNFAGPDLLHDPSDKTFYDDPETCYTMMDKPMKNWDEKRKEWLFHHPSFAAGATEKI|
PHD sec  |                              EE         HHHHHHHHHHH  HHHHHHHEEE|
Rel sec  |32245589975756997664789711224336633562688999885346146553215 7|

         ....,....19...,....20...,....21...,....22...,....23...,....24
AA       |AAMLAHPEVEWIWWVDSDAIFTDMEFKLPLWRYKDHNLVIHGWEELVKTEHSWTGLNAGV|
PHD sec  |HHHHH   EEEEEEE       HHH      EE    EEEE    EEE    E       |
Rel sec  |99961999579987336411122103455220148851552521011133212332221 1|

         ....,....25...,....26...,....27...,....28...,....29...,....30
AA       |FLMRNCQWSLDFMDVWASMGPNSPEYEKWGERLRETFKTKVVRDSDDQTALAYLIAMGED|
PHD sec  |EEEE  HHHHHHHHHHHHH     HHHHHHHHHHHHH  EEEEE  HHHHHHHHH    |
Rel sec  |43431113255544374618999387887899999633323652817999999998265 3|

         ....,....31...,....32...,....33...,....34...,....35...,....36
AA       |KWTKKIYMENEYYFEGYWLEISKMYDKMGERYDEIEKRVEGLRRRHAEKVSERYGEMREE|
PHD sec  |    EEEEEEEEEE  EEEEEEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH|
Rel sec  |11321332342342222546415577665779999999999999999999999999999|

         ....,....37...,....38...,....39...,....40...,....41...,....42
AA       |YVKNLGDMRRPFITHFTGCQPCNGHHNPIYAADDCWNGMERALNFADNQVLRKFGFIHPN|
PHD sec  |HHHHHH    EEEEE                  HHHHHHHHHHHHHH HHHHHHH      |
Rel sec  |99987337860466513568687788875554247789999999842699999612342 0|

         ....,....43...,....44...,....45...,....46...,....47...,....48
AA       |LLDKSVSPLPFGYPAASP|
PHD sec  |                  |
Rel sec  |14346997888999999|
```

Fig. 6A

10/11

```
          ....,....1....,....2....,....3....,....4....,....5....,....6
AA        |ATKFGSKNKSSPWLSNGCIFLLGAMSALLMIWGLNSFIAPIPNSNPKFNSFTTKLKSLNF|
PHD htm   |                    TTTTTTTTTTTTTTTTTTTTT
          ....,....7....,....8....,....9....,....10...,....11...,....12
AA        |TTNTNFAGPDLLHDPSDKTFYDDPETCYTMMDKPMKNWDEKRKEWLFHHPSFAAGATEKI|
PHD htm   |                                                             |
          ....,....13...,....14...,....15...,....16...,....17...,....18
AA        |LVITGSQPTKCDNPIGDHLLLRFYKNKVDYCRIHNHDIIYNNALLHPKMDSYWAKYPMVR|
PHD htm   |                                                             |
          ....,....19...,....20...,....21...,....22...,....23...,....24
AA        |AAMLAHPEVEWIWWVDSDAIFTDMEFKLPLWRYKDHNLVIHGWEELVKTEHSWTGLNAGV|
PHD htm   |                                                             |
          ....,....25...,....26...,....27...,....28...,....29...,....30
AA        |FLMRNCQWSLDFMDVWASMGPNSPEYEKWGERLRETFKTKVVRDSDDQTALAYLIAMGED|
PHD htm   |                                                             |
          ....,....31...,....32...,....33...,....34...,....35...,....36
AA        |KWTKKIYMENEYYFEGYWLEISKMYDKMGERYDEIEKRVEGLRRRHAEKVSERYGEMREE|
PHD htm   |                                                             |
          ....,....37...,....38...,....39...,....40...,....41...,....42
AA        |YVKNLGDMRRPFITHFTGCQPCNGHHNPIYAADDCWNGMERALNFADNQVLRKFGFIHPN|
PHD htm   |                                                             |
          ....,....43...,....44...,....45...,....46...,....47...,....48
AA        |LLDKSVSPLPFGYPAASP|
PHD htm   |                  |
```

Fig. 6B

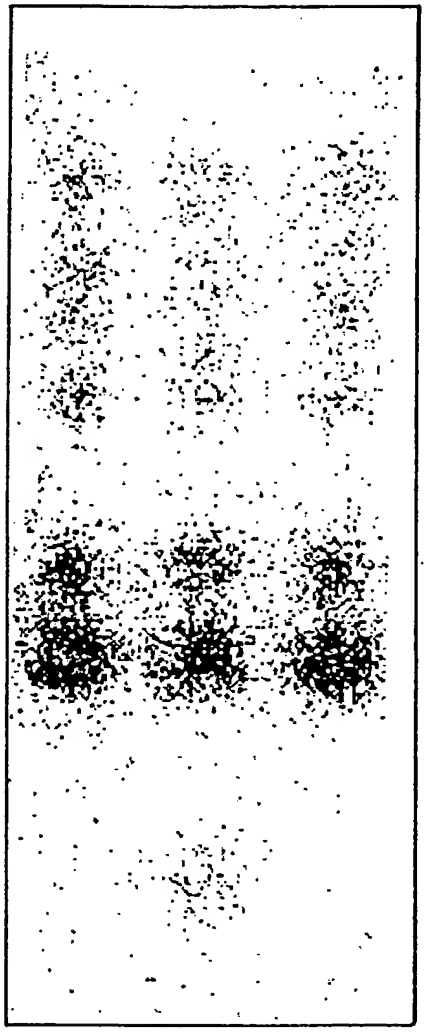

Fig. 7

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 9/10, A01H 5/00, A23C 9/00, C07K 16/40, C12N 5/04, C12N 15/82, G01N 33/53

A3

(11) International Publication Number: WO 99/60103

(43) International Publication Date: 25 November 1999 (25.11.1999)

(21) International Application Number: PCT/GB99/01610

(22) International Filing Date: 21 May 1999 (21.05.1999)

(30) Priority Data:
9810997.8 21 May 1998 (21.05.1998) GB

(60) Parent Application or Grant
UNILEVER PLC [/]; (). UNILEVER N.V. [/]; (). HINDUSTAN LEVER LIMITED [/]; (). GIDLEY, Michael, John [/]; (). CHENGAPPA, Sumant [/]; (). REID, John, Spence, Grant [/]; (). EDWARDS, Mary, Elizabeth [/]; (). DICKSON, Cathryn, Anne [/]; (). GIDLEY, Michael, John [/]; (). CHENGAPPA, Sumant [/]; (). REID, John, Spence, Grant [/]; (). EDWARDS, Mary, Elizabeth [/]; (). DICKSON, Cathryn, Anne [/]; (). KREMER, Simon, M. ; ().

Published

(54) Title: GALACTOSYLTRANSFERASE FROM PLANTS INVOLVED IN GALACTOMANNAN BIOSYNTHRSIS

(54) Titre: GENES ET ENZYMESERASE FROM PLANTS INVOLVED IN GALACTOMANNAN BIOSYNTHRSIS

(57) Abstract

Provided are genes encoding enzymes (e.g. galactosyltransferases from fenugreek and guar) which have a role in the biosynthesis of complex non-cellulosic cell wall polysaccharides such as galactomannan. Variants and other products based on the genes are also provided, as are antibodies to the enzymes, plus also methods of isolating or preparing any of these. Also disclosed are vectors and other methods and materials which may be used for cloning the genes (or related nucleic acids e.g. anti-sense versions) into hosts such as transgenic plants having modified polysaccharides. Further disclosed are products, compositions and methods employing these plants and polysaccharides.

(57) Abrégé

L'invention porte sur des gènes codant des enzymes (telles que des galactosyltransférases issues du fenugrec et du guar) qui jouent un rôle dans la biosynthèse de polysaccharides complexes à parois cellulaires non cellulosiques tels que des galactomannans. L'invention porte également sur des variantes et autres produits à base de gènes, ainsi que sur des anticorps contre les enzymes, et sur des procédés d'isolation ou de préparation de l'un quelconque de ces produits. L'invention porte sur des vecteurs et autres procédés et matériaux qui peuvent être utilisés pour cloner les gènes (ou des acides nucléiques apparentés tels que des versions antisens) dans des hôtes tels que des plantes transgéniques possédant des polysaccharides modifiés. L'invention porte en outre sur des produits, des compositions et des procédés utilisant ces plantes et ces polysaccharides.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/10, 15/82, G01N 33/53, C12N 5/04, A01H 5/00, C07K 16/40, A23C 9/00	A3	**(11) International Publication Number:** WO 99/60103 **(43) International Publication Date:** 25 November 1999 (25.11.99)

(21) International Application Number: PCT/GB99/01610

(22) International Filing Date: 21 May 1999 (21.05.99)

(30) Priority Data:
9810997.8 21 May 1998 (21.05.98) GB

(71) Applicant *(for AE AU BB CA CY GB GD GH GM IE IL KE LC LK LS MN MW NZ SD SG SL SZ TT UG ZA ZW only)*: UNILEVER PLC [GB/GB]; Unilever House, Blackfriars, London EC4P 4BQ (GB).

(71) Applicant *(for all designated States except AE AU BB CA CY GB GD GH GM IE IL IN KE LC LK LS MN MW NZ SD SG SL SZ TT UG US ZA ZW)*: UNILEVER N.V. [NL/NL]; Weena 455, NL-3013 AL Rotterdam (NL).

(71) Applicant *(for IN only)*: HINDUSTAN LEVER LIMITED [IN/IN]; Hindustan Lever House, 165/166 Backbay Reclamation, Mumbai 400 020, Maharashtra (IN).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GIDLEY, Michael, John [GB/GB]; Unilever Research, Colworth Laboratory, Colworth House, Sharnbrook, Bedford MK44 1LQ (GB). CHENGAPPA, Sumant [IN/GB]; Unilever Research, Colworth Laboratory, Colworth House, Sharnbrook, Bedford MK44 1LQ (GB). REID, John, Spence, Grant [GB/GB]; University of Sterling, Dept. of Biological Sciences, Stirling FK9 4LA (GB). EDWARDS, Mary, Elizabeth [GB/GB]; University of Sterling, Dept. of Biological Sciences, Stirling FK9 4LA (GB). DICKSON, Cathryn, Anne [GB/GB]; University of Sterling, Dept. of Biological Sciences, Stirling FK9 4LA (GB).

(74) Agents: KREMER, Simon, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
2 March 2000 (02.03.00)

(54) Title: GALACTOSYLTRANSFERASE FROM PLANTS INVOLVED IN GALACTOMANNAN BIOSYNTHESIS

(57) Abstract

Provided are genes encoding enzymes (e.g. galactosyltransferases from fenugreek and guar) which have a role in the biosynthesis of complex non-cellulosic cell wall polysaccharides such as galactomannan. Variants and other products based on the genes are also provided, as are antibodies to the enzymes, plus also methods of isolating or preparing any of these. Also disclosed are vectors and other methods and materials which may be used for cloning the genes (or related nucleic acids e.g. anti-sense versions) into hosts such as transgenic plants having modified polysaccharides. Further disclosed are products, compositions and methods employing these plants and polysaccharides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

Inte ional Application No
PCT/GB 99/01610

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/10 C12N15/82 G01N33/53 C12N5/04 A01H5/00
C07K16/40 A23C9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL EBI Accession no. AB010070, 9 January 1998 (1998-01-09) NAKAMURA Y. ET AL.: "Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone" XP002125788 see Sequence abstract — -/—	1-17, 24-30, 32,34,35

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 December 1999	11/01/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Chakravarty, A

2

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Intl cnal Application No
PCT/GB 99/01610

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL EBI Accession no. AC004786, 4 June 1998 (1998-06-04) ROUNSLEY S.D. ET AL.: "Arabidopsis thaliana chromoscne II BAC T20k9 genomic sequence" XP002125757 see Sequence abstract	1-17, 24-30, 32,34,35
P,X	WO 99 01558 A (UNIV CAMBRIDGE TECH ;DUPREE PAUL (GB); MOEGELSVANG SOEREN (GB)) 14 January 1999 (1999-01-14) claim 8; figure 4C claims 39,48	1-3,6-9, 11-68
A	WO 97 20937 A (DANISCO ;JOERSBOE MORTEN (DK); BRUNSTEDT JANNE (DK); PETERSEN STEE) 12 June 1997 (1997-06-12) cited in the application the whole document	
A	EDWARDS M ET AL: "CONTROL OF MANNOSE/GALACTOSE RATIO DURING GALACTOMANNAN FORMATION IN DEVELOPING LEGUME SEEDS" PLANTA,DE,SPRINGER VERLAG, vol. 187, no. 1, 1992, page 67-74 XP000654833 ISSN: 0032-0935 cited in the application	
A	EDWARDS ET AL: "Biosynthesis of legume-seed galactomannans in vitro" PLANTA,DE,SPRINGER VERLAG, no. 178, 1989, page 41-51 XP002076760 ISSN: 0032-0935 cited in the application	

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter[illegible]nal Application No
PCT/GB 99/01610

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9901558 A	14-01-1999	AU 8226598 A	25-01-1999
WO 9720937 A	12-06-1997	AU 1193997 A	27-06-1997
		CN 1207772 A	10-02-1999
		EP 0866869 A	30-09-1998